

Beta

Selecting a {"no":"SQL"}
database from a menu of
over 200 products

Akmal B. Chaudhri

(艾克摩 曹理)

Why it's important



Half of the “NoSQL” databases and “big data” technologies that are hot buzzwords won’t be around in 15 years.

-- Michael O. Church

Agenda



In a packed program ...

- Introduction
- Market analysis
- NoSQL
- Security and vulnerability
- Polyglot persistence
- Benchmarks and performance
- BI/Analytics
- NoSQL alternatives
- Summary
- Resources

In a packed program ...

- Introduction
- Market analysis
- NoSQL
- Security and vulnerability
- Polyglot persistence
- Benchmarks and performance
- BI/Analytics
- NoSQL alternatives
- Summary
- Resources

In a packed program ...

- Introduction
- Market analysis
- **NoSQL**
- Security and vulnerability
- Polyglot persistence
- Benchmarks and performance
- BI/Analytics
- NoSQL alternatives
- Summary
- Resources

In a packed program ...

- Introduction
- Market analysis
- NoSQL
- **Security and vulnerability**
- **Polyglot persistence**
- **Benchmarks and performance**
- **BI/Analytics**
- **NoSQL alternatives**
- Summary
- Resources

In a packed program

- Introduction
- Market analysis
- NoSQL
- Security and vulnerability
- Polyglot persistence
- Benchmarks and performance
- BI/Analytics
- NoSQL alternatives
- **Summary**
- **Resources**

Introduction

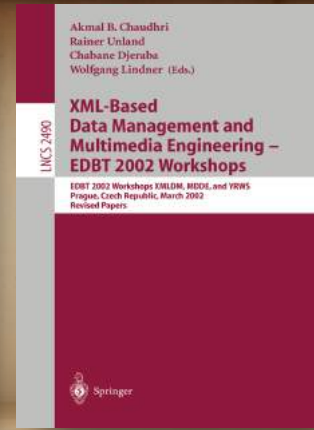
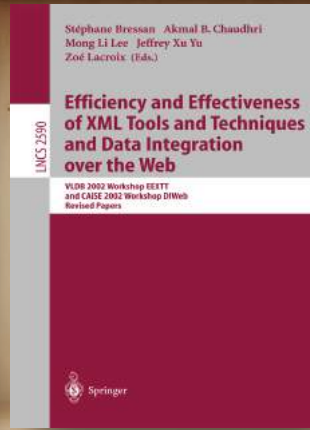
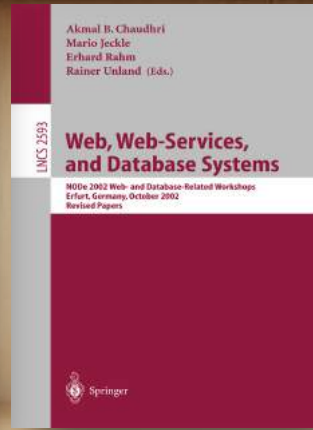
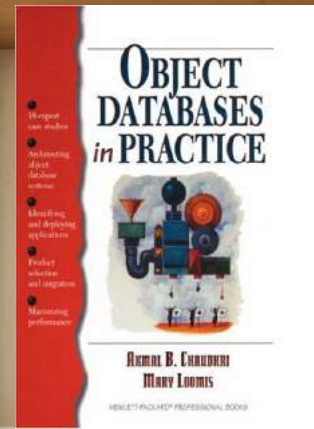
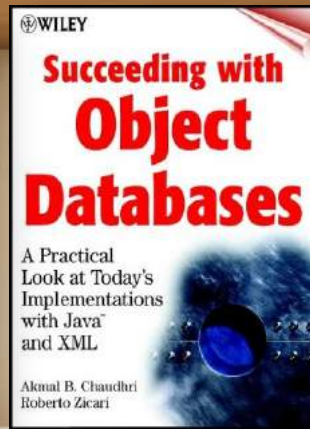
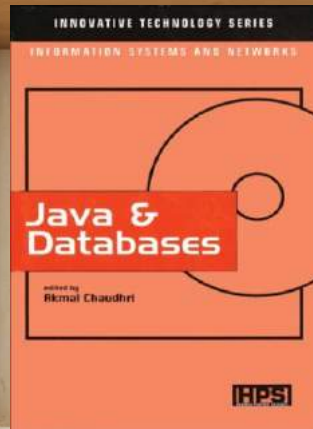
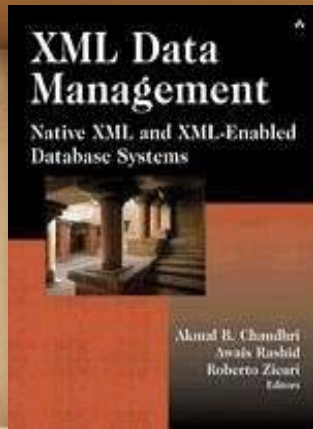


My background

- ~25 years experience in IT
 - Developer (Reuters)
 - Academic (City University)
 - Consultant (Logica)
 - Technical Architect (CA)
 - Senior Architect (Informix)
 - Senior IT Specialist (IBM)
 - TI (Hortonworks)
 - SA (DataStax)
- Worked with various technologies
 - Programming languages
 - IDE
 - Database Systems
- Client-facing roles
 - Developers
 - Senior executives
 - Journalists
- Broad industry experience
- Community outreach
- University relations
- 10 books, many presentations

Full disclosure

- Worked for
 - DataStax
- Consulted for
 - MongoDB
 - VoltDB



Old Java user group



- London JSIG was amongst the top 25 Java User Groups in the world, as voted by members

History

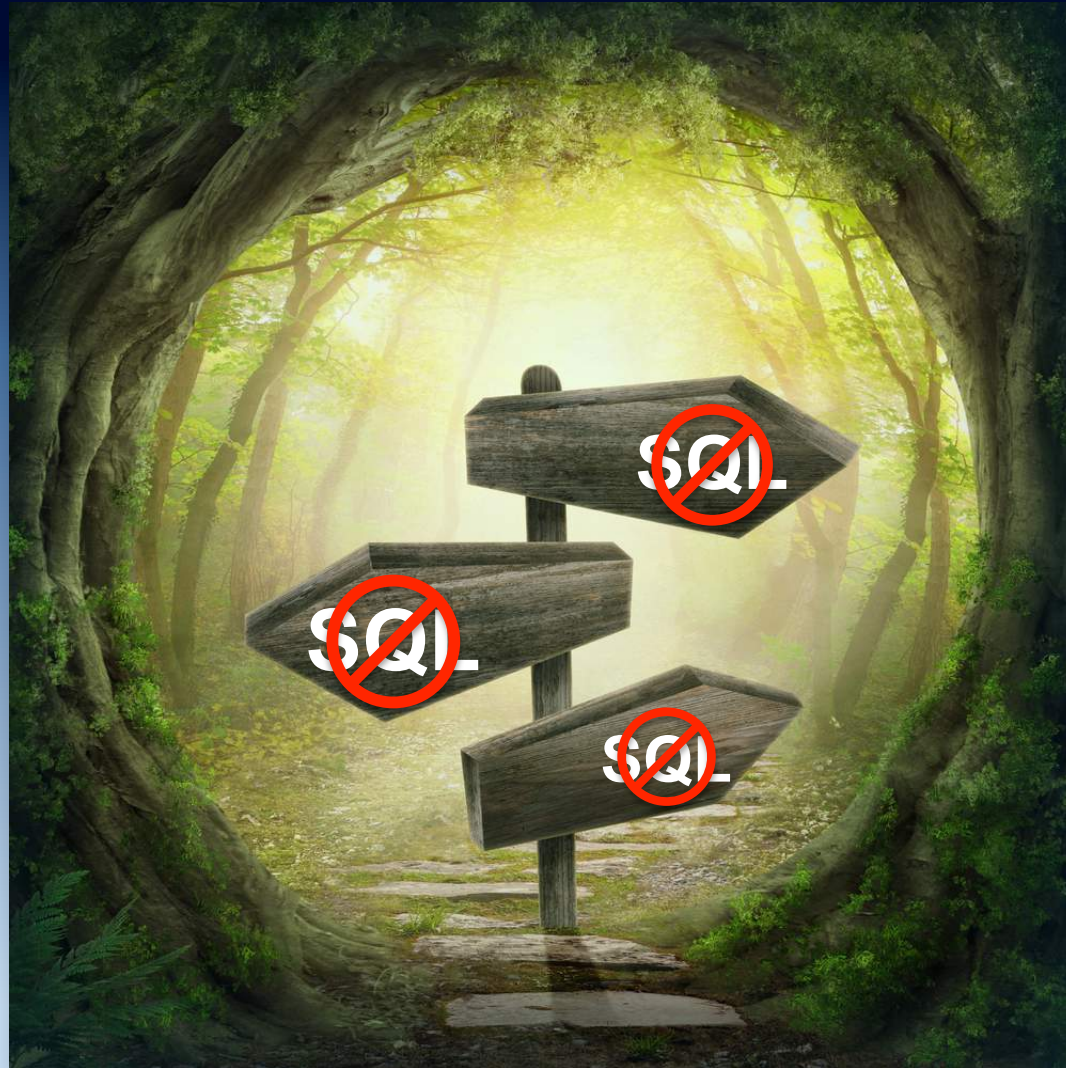


Have you run into limitations with traditional relational databases? Don't mind trading a query language for scalability? Or perhaps you just like shiny new things to try out? Either way this meetup is for you.

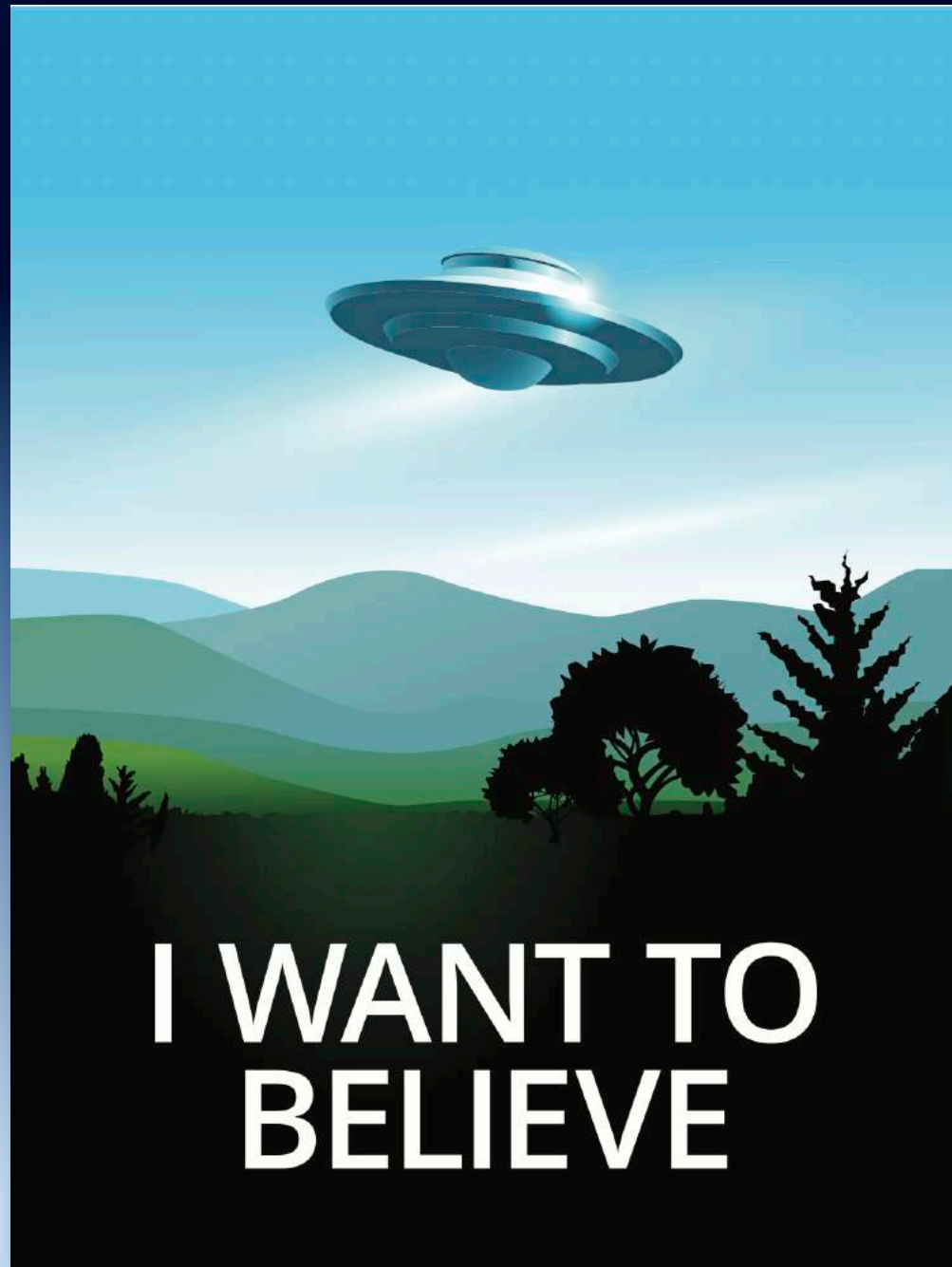
Join us in figuring out why these new fangled Dynamo clones and BigTables have become so popular lately.



Your path leads to NoSQL?

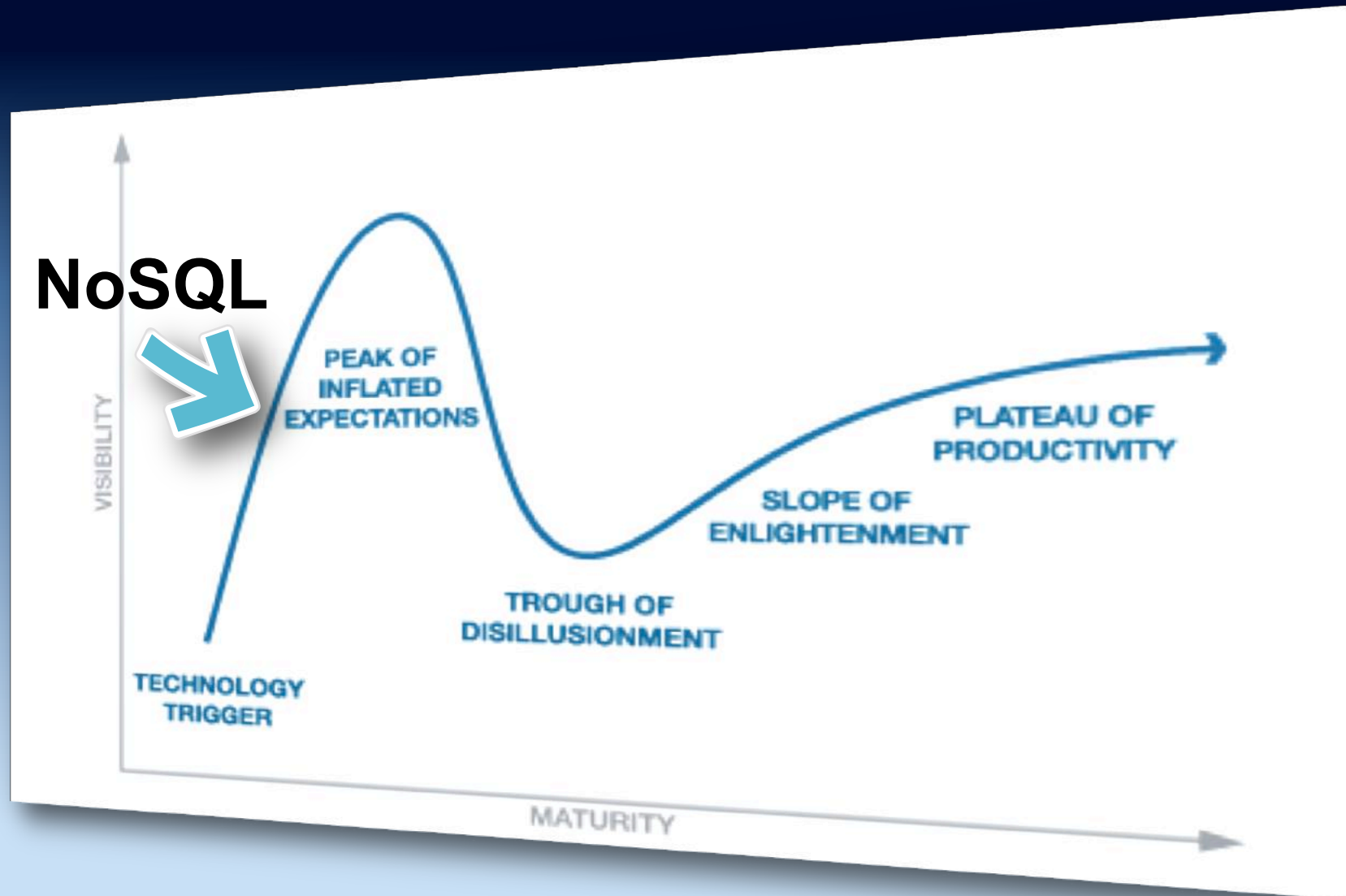


Source: Shutterstock Image ID 159183185



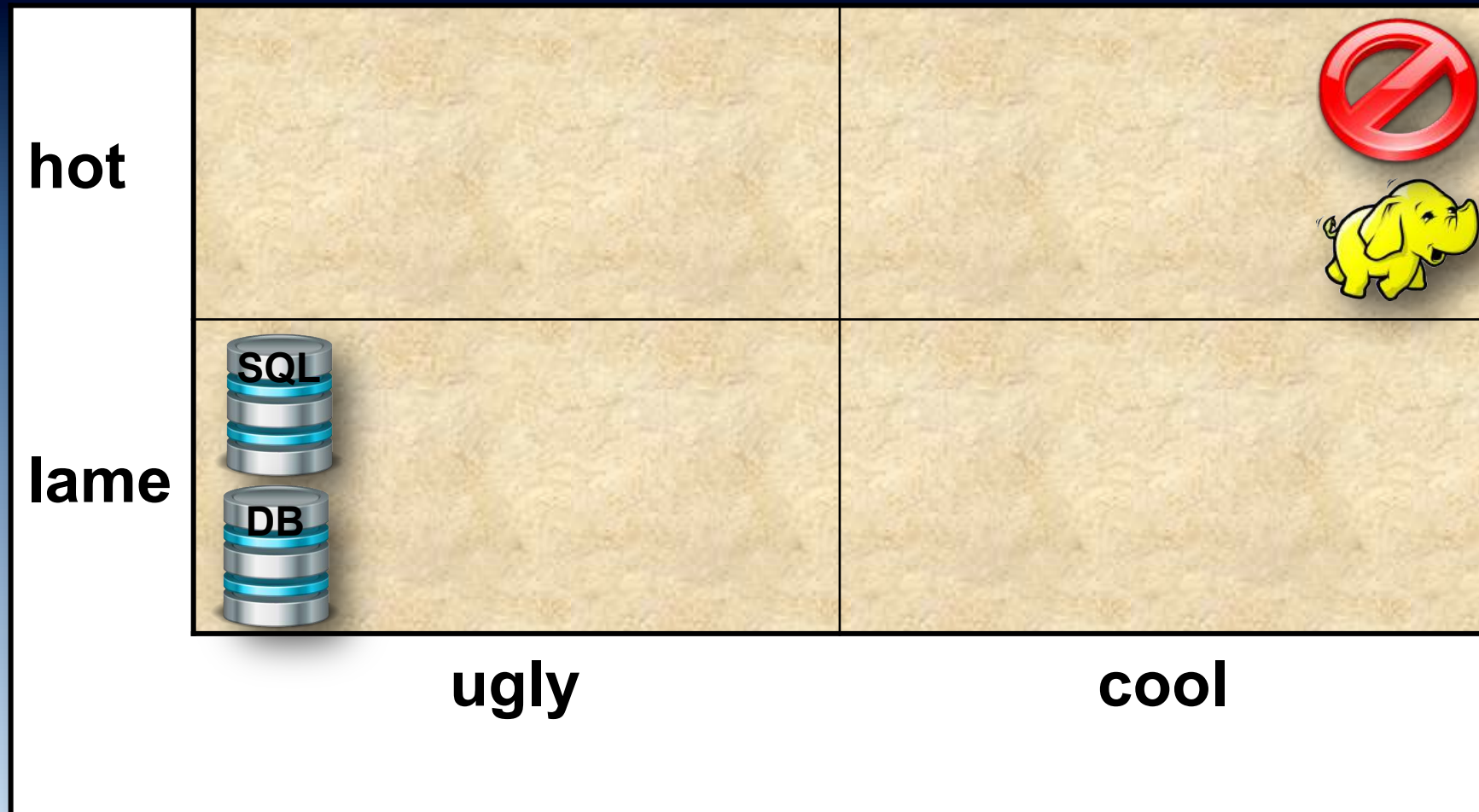
Source: Shutterstock Image ID 99862922

Gartner hype curve



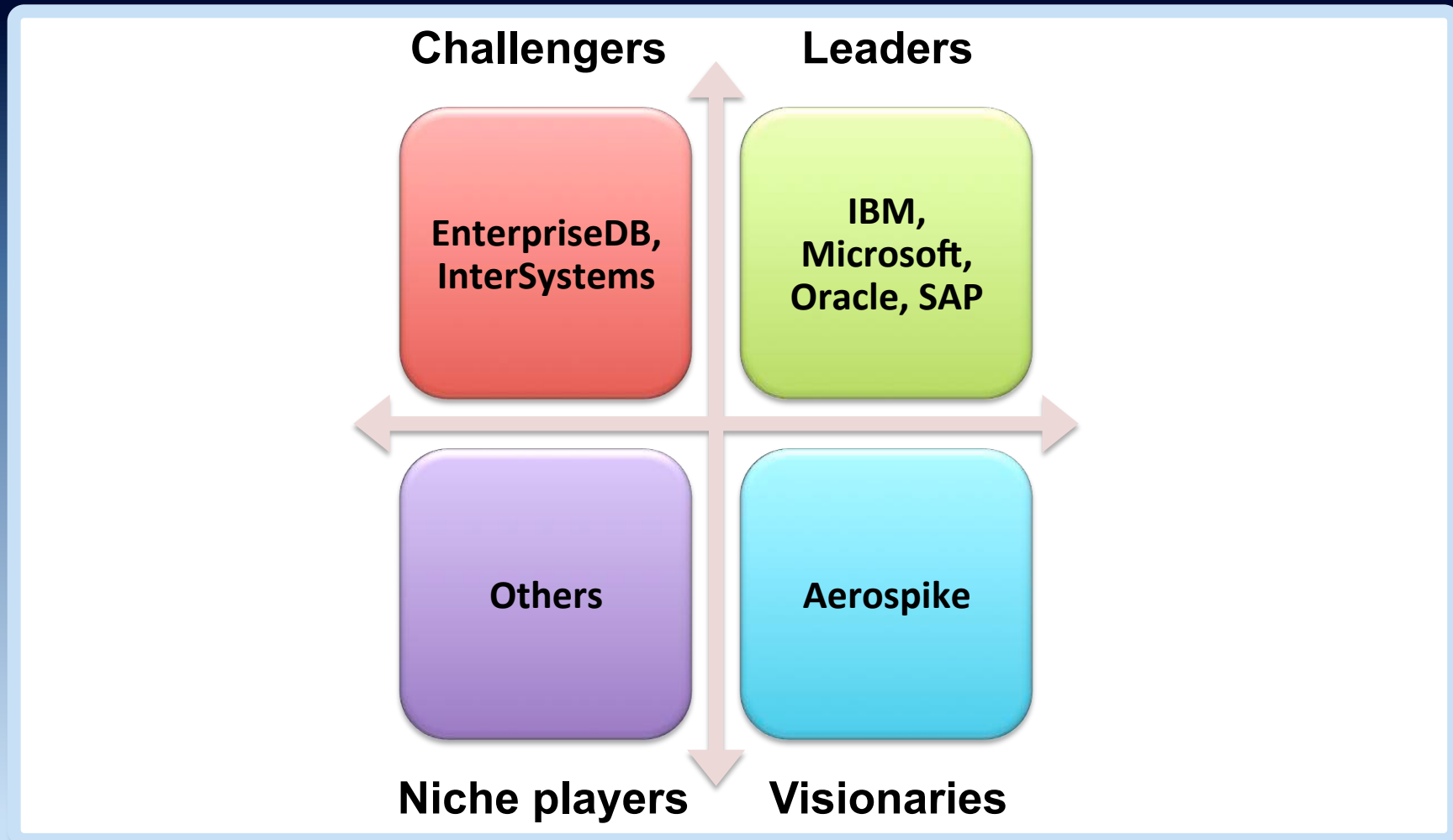


Magic quadrant



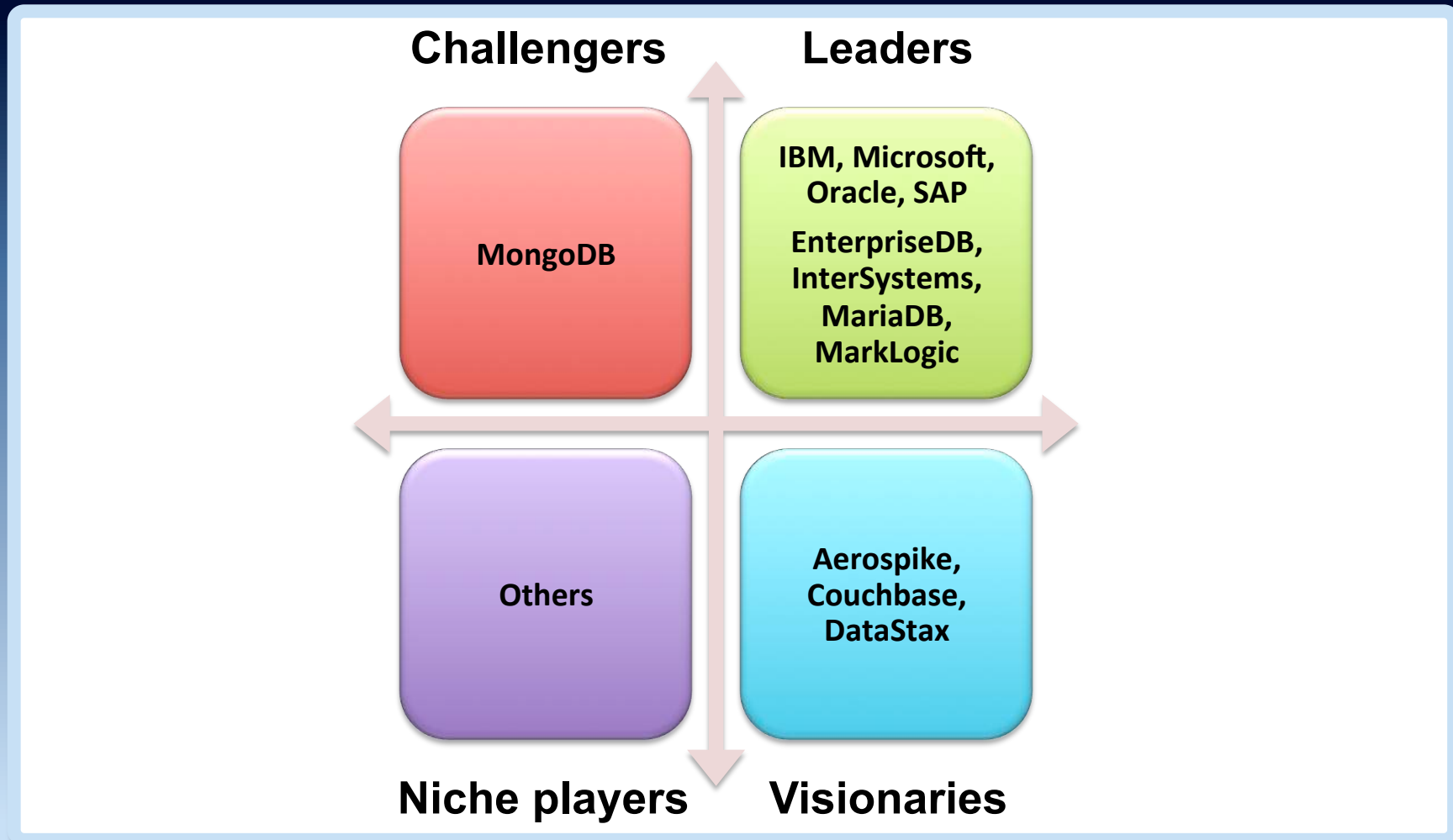
Source: After “say No! No! and No! (=NoSQL Parody)” Jens Dittrich (2013)

Magic quadrant 2013



Source: "Magic Quadrant for Operational Database Management Systems" Gartner (21 October 2013)

Magic quadrant 2014



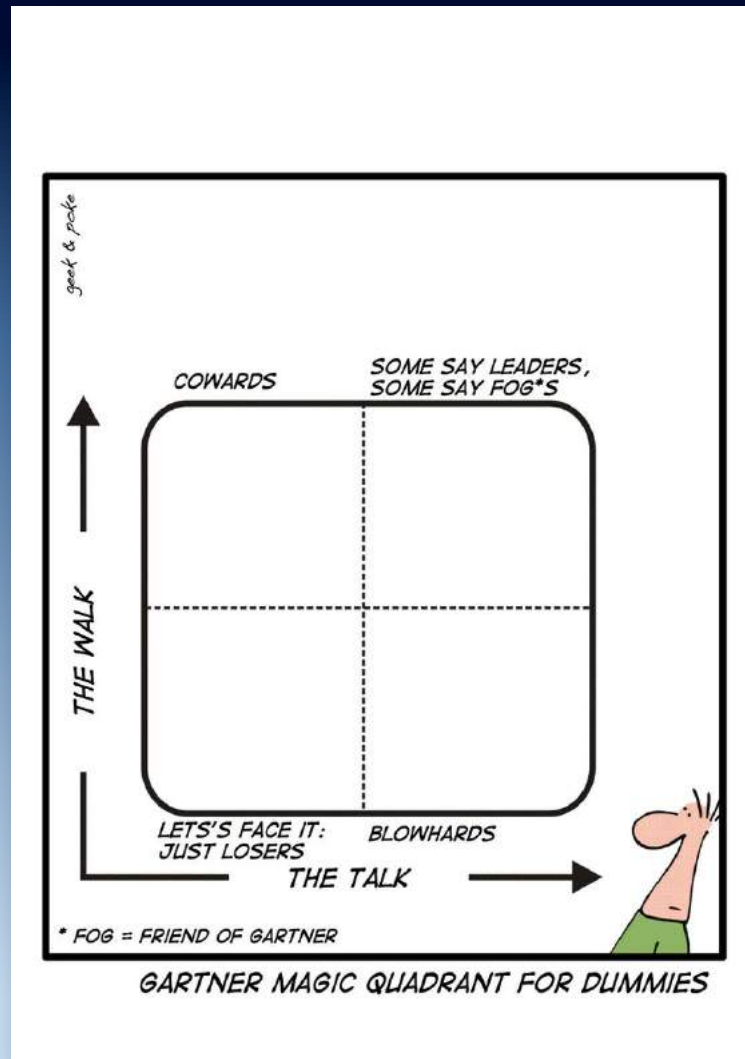
Source: "Magic Quadrant for Operational Database Management Systems" Gartner (16 October 2014)

Magic quadrant 2015



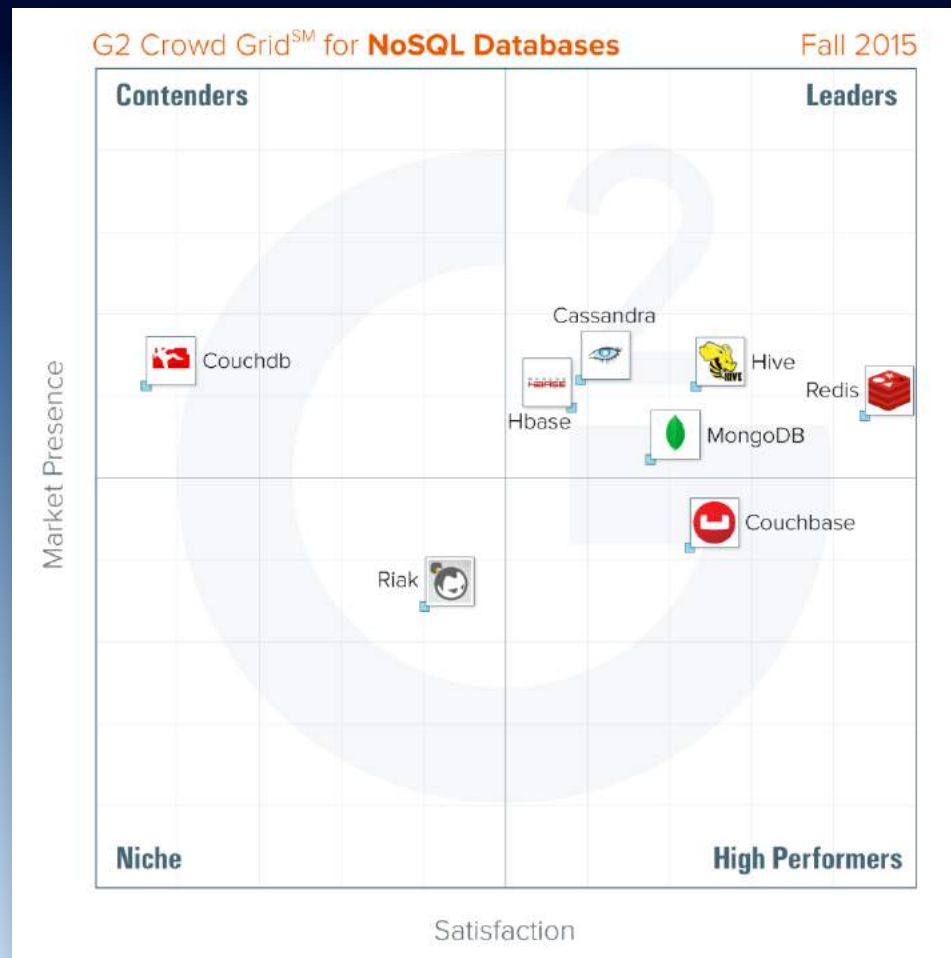
Source: "Magic Quadrant for Operational Database Management Systems" Gartner (12 October 2015)

Magic quadrant for dummies



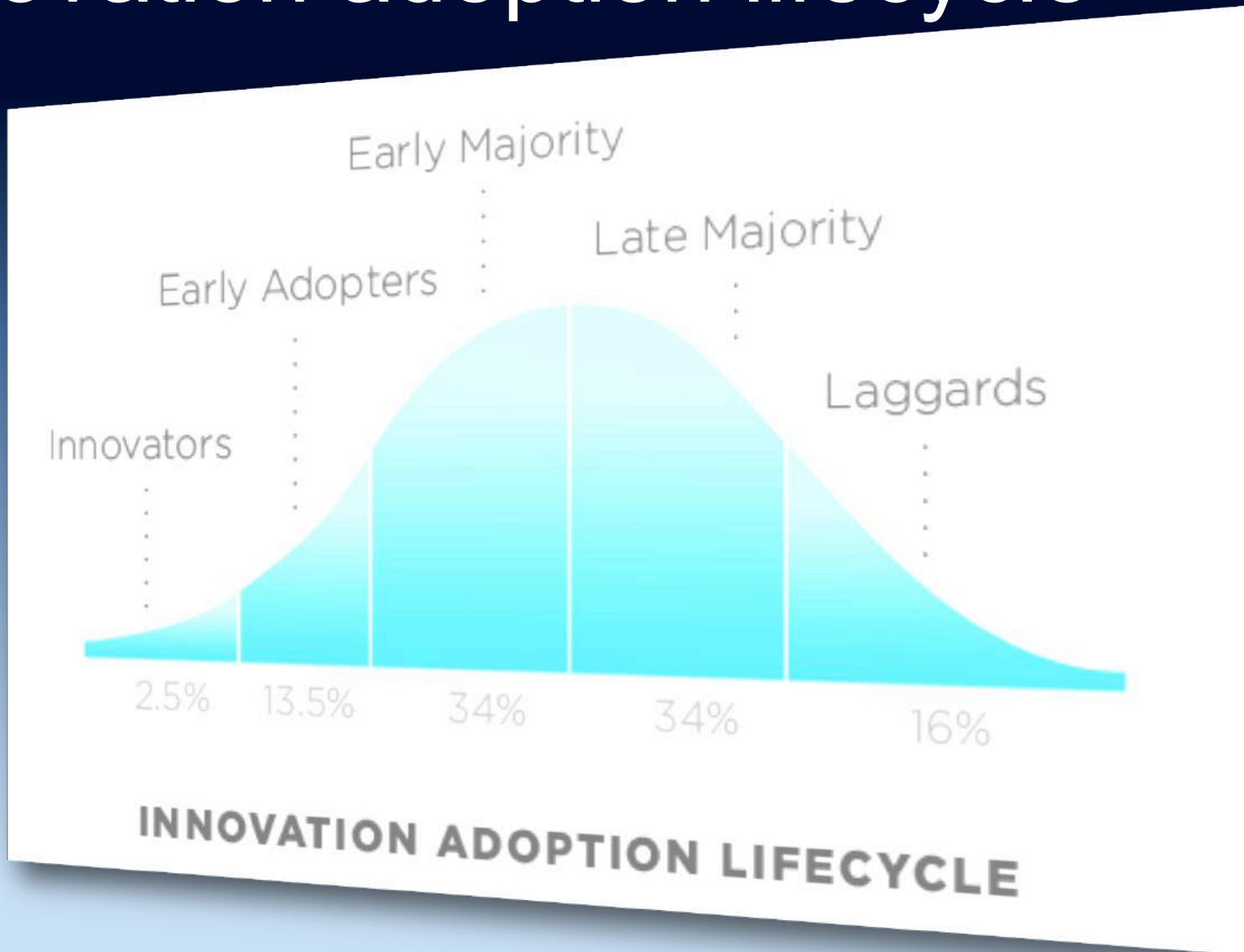
Source: Oliver Widder, used with permission

G2 Crowd Grid for NoSQL



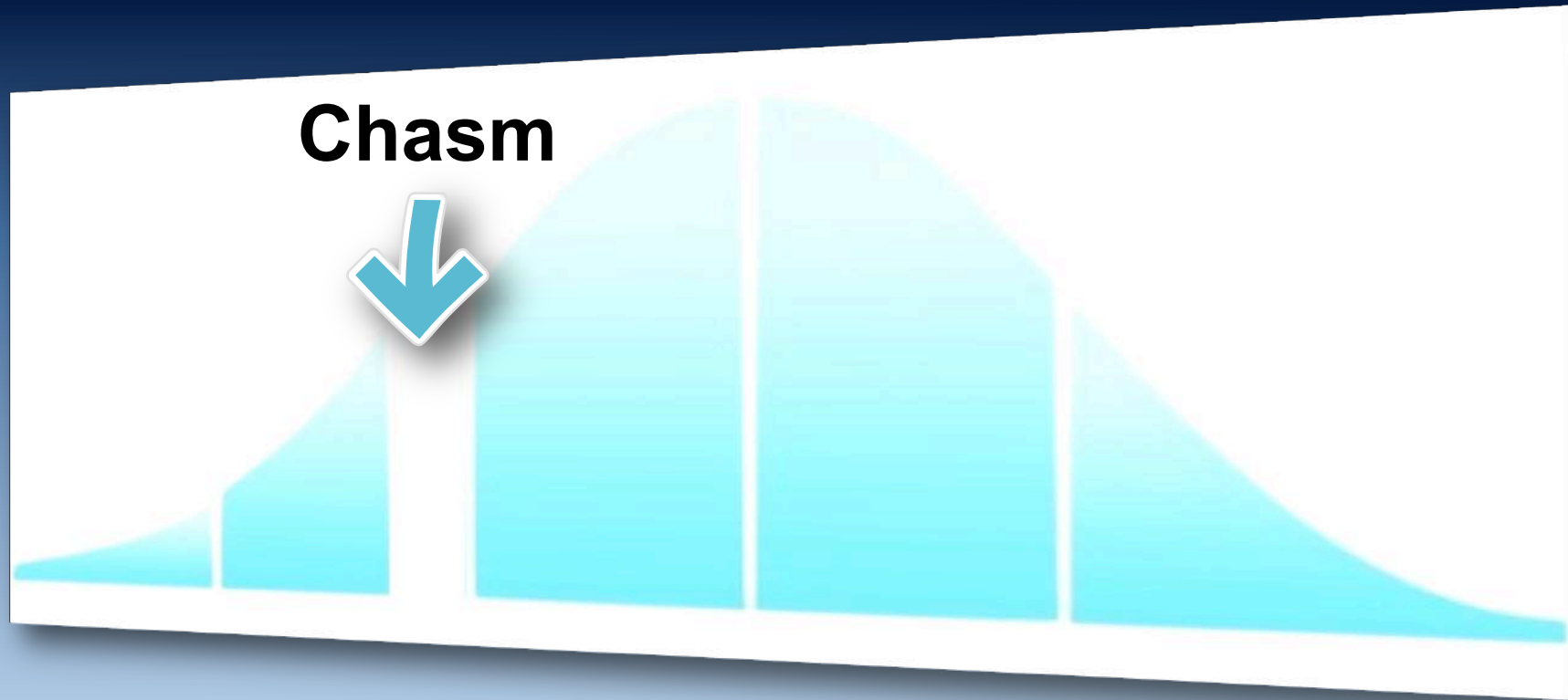
Source: G2 Crowd, used with permission

Innovation adoption lifecycle

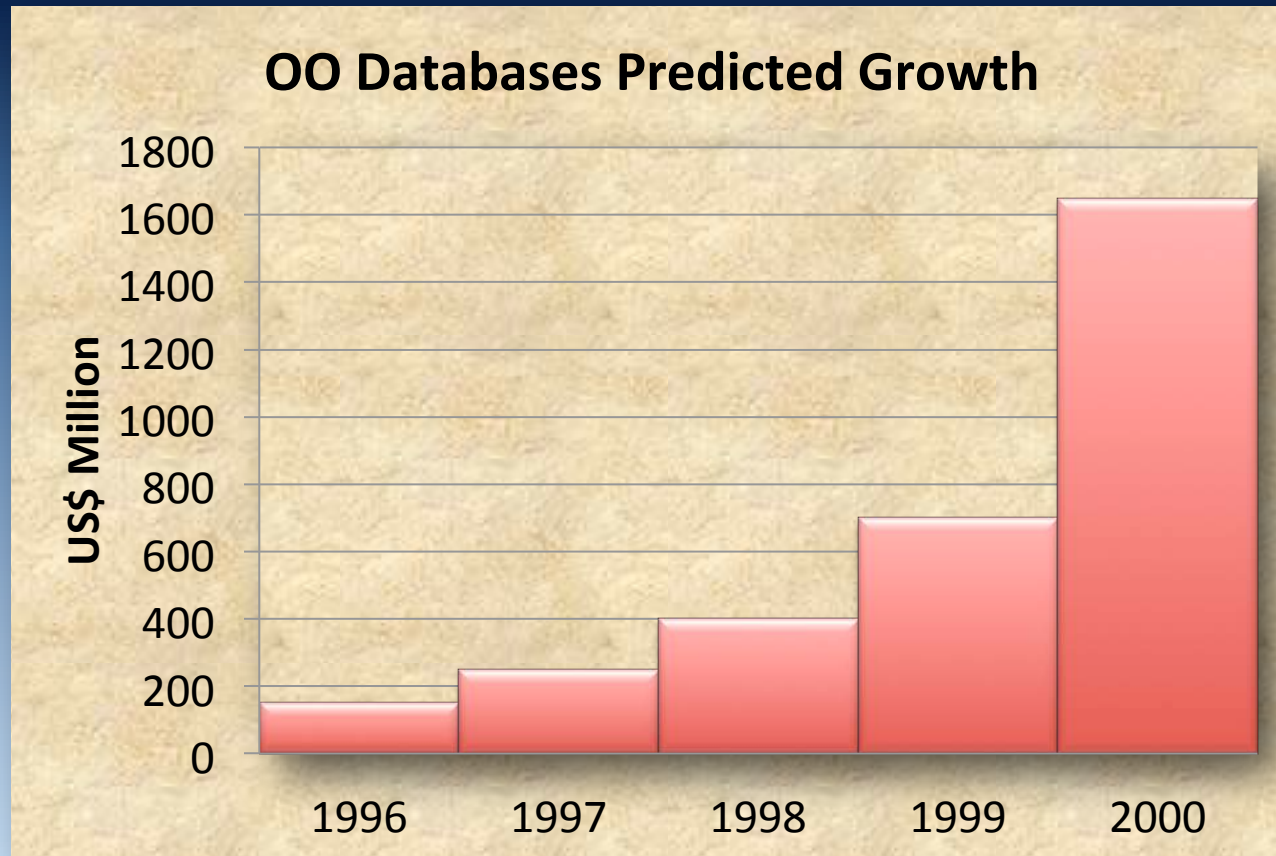


Source: http://en.wikipedia.org/wiki/Technology_adoption_lifecycle

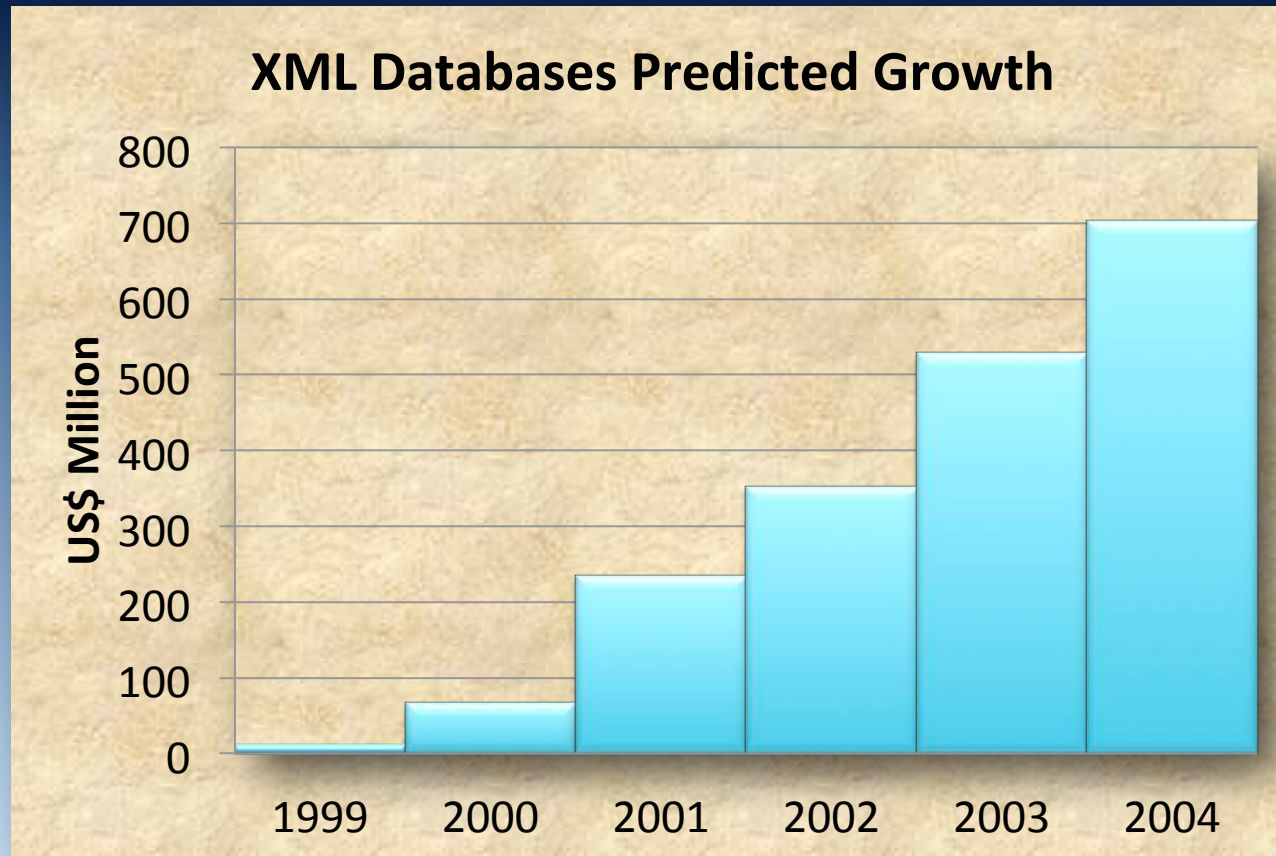
Crossing the chasm



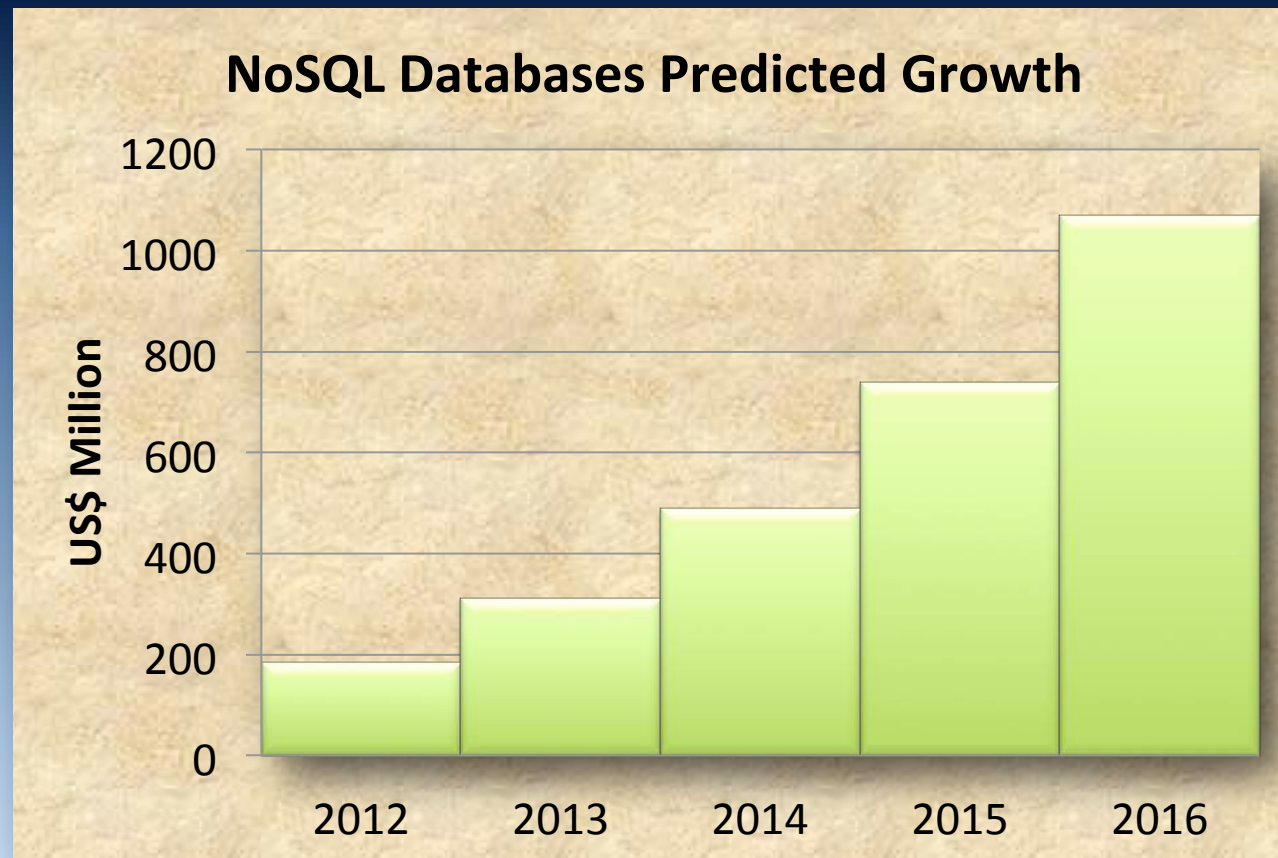
1990s



2000s



Today



The way developers really think



OO vs. Relational



Source: Inspired by comments from Esther Dyson during the 1990s

XML vs. Relational



Source: Inspired by "Tamino - What is it good for?" Curtis Pew (2003)

NoSQL vs. Relational

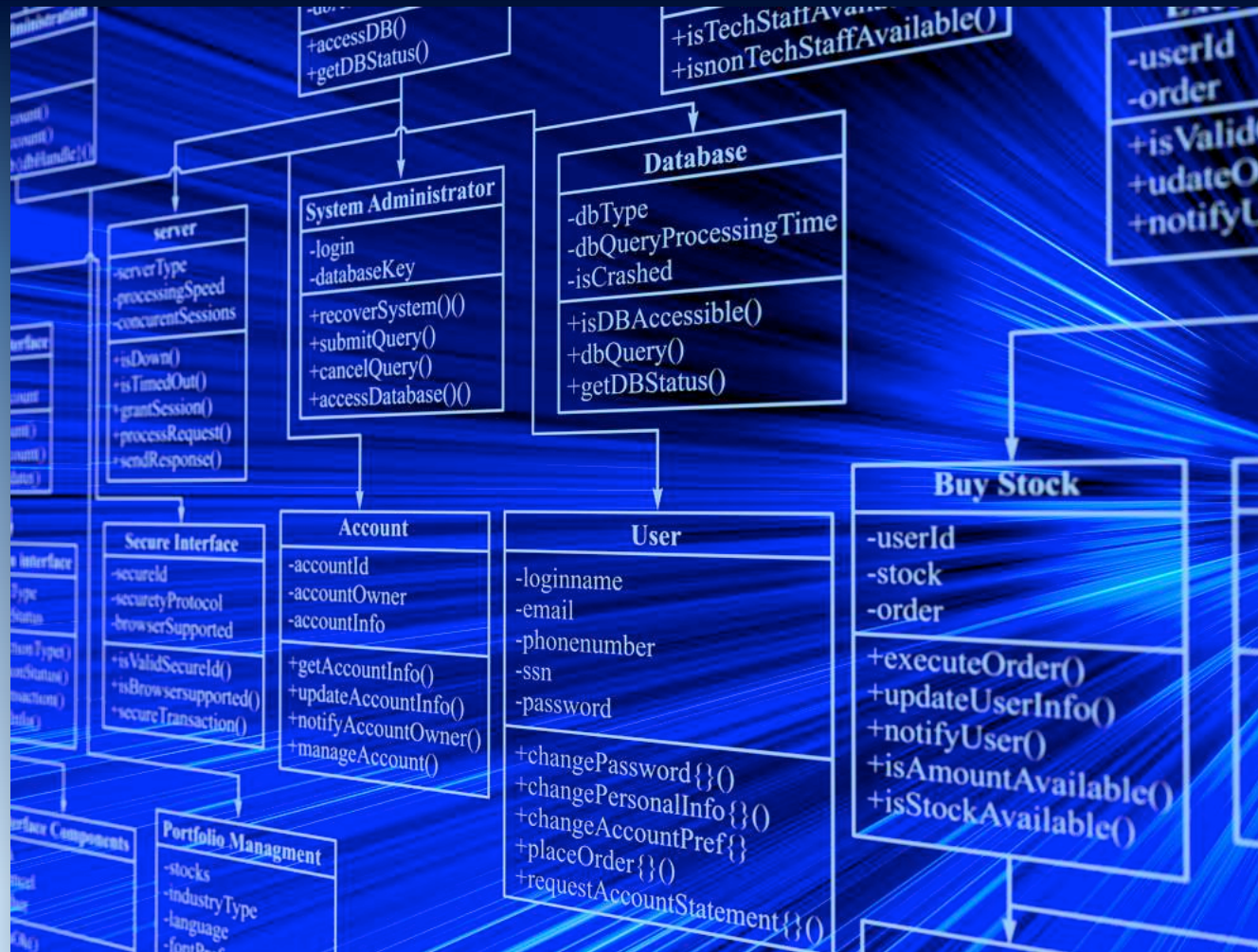


Source: Inspired by “Data Management for Interactive Applications” Couchbase (12 June 2013) and “MongoDB and the OpEx Business Plan” MongoDB (9 July 2013)

But ...

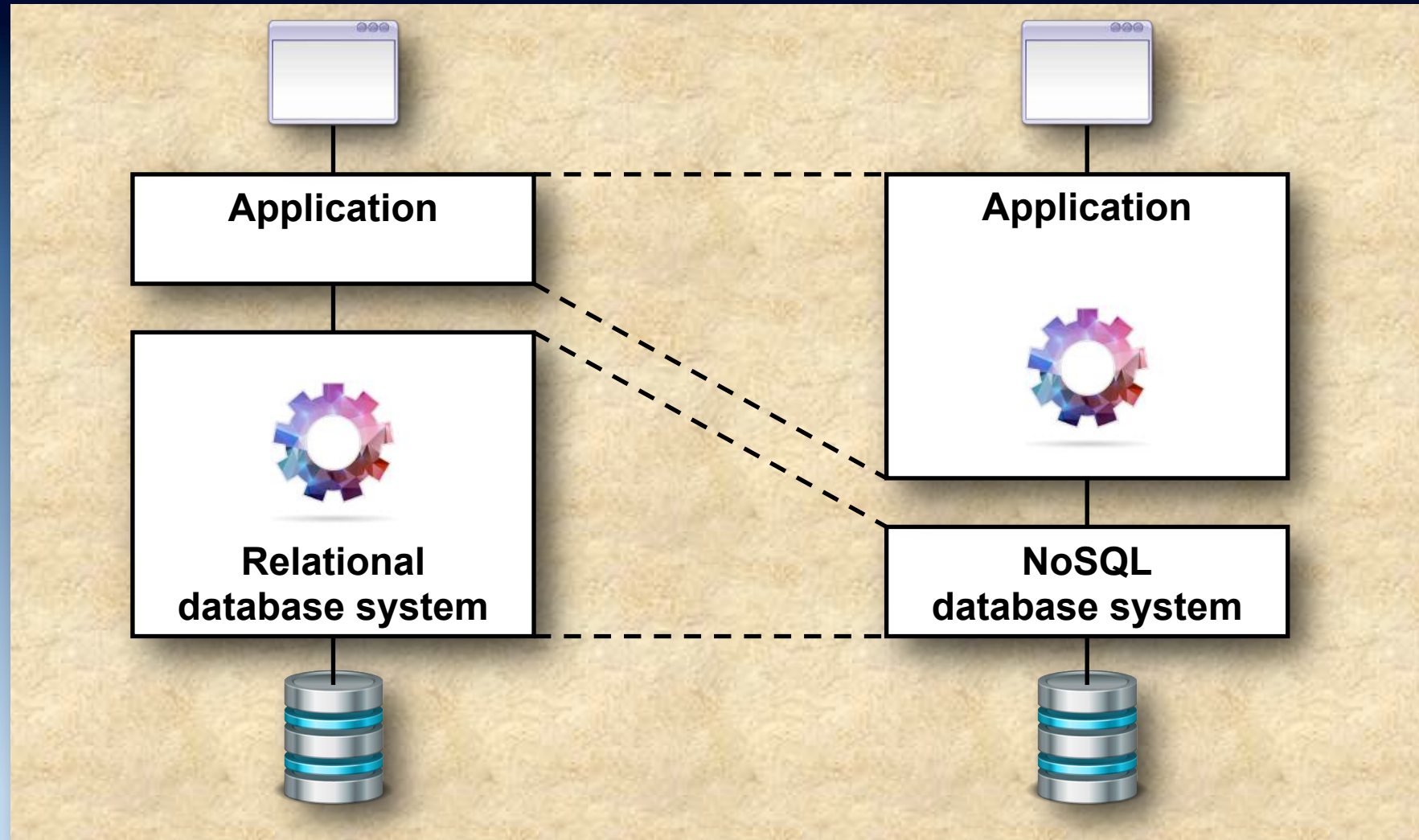


Relational flexibility



Source: Shutterstock Image ID 73381360

Welcome to 1985 ...



Source: After "NoSQL and the responsibility shift" Denshade (14 March 2015)

Welcome to 1985



NoSQL-only solutions also only store data. They don't process it. Data must be brought to the application for analysis. The application (and hence each individual application developer) is responsible for efficiently accessing data, implementing business rules, and for data consistency.

-- Pierre Fricke

“MongoDB is web scale”



It may surprise you that there are a handful of high-profile websites still using relational databases and in particular MySQL.



NoSQL is developer-friendly

Other Stakeholders



Developers



But ...



Riak ... We're talking about nearly a year of learning.^[1]

Things I wish I knew about MongoDB a year ago^[2]

I am learning Cassandra. It is not easy.^[3]

[1] <http://production-scale.com/blog/2011/11/20/building-an-application-upon-riak-part-1.html>

[2] <http://snmaynard.com/2012/10/17/things-i-wish-i-knew-about-mongodb-a-year-ago/>

[3] <http://planetcassandra.org/blog/post/datastax-java-driver-for-apache-cassandra>

And ...



... it takes 1-3 years to get an enterprise application onto a new data platform like Cassandra ... Cassandra requires a complete re-thinking of the data model which many find challenging.

-- Shanti Subramanyam

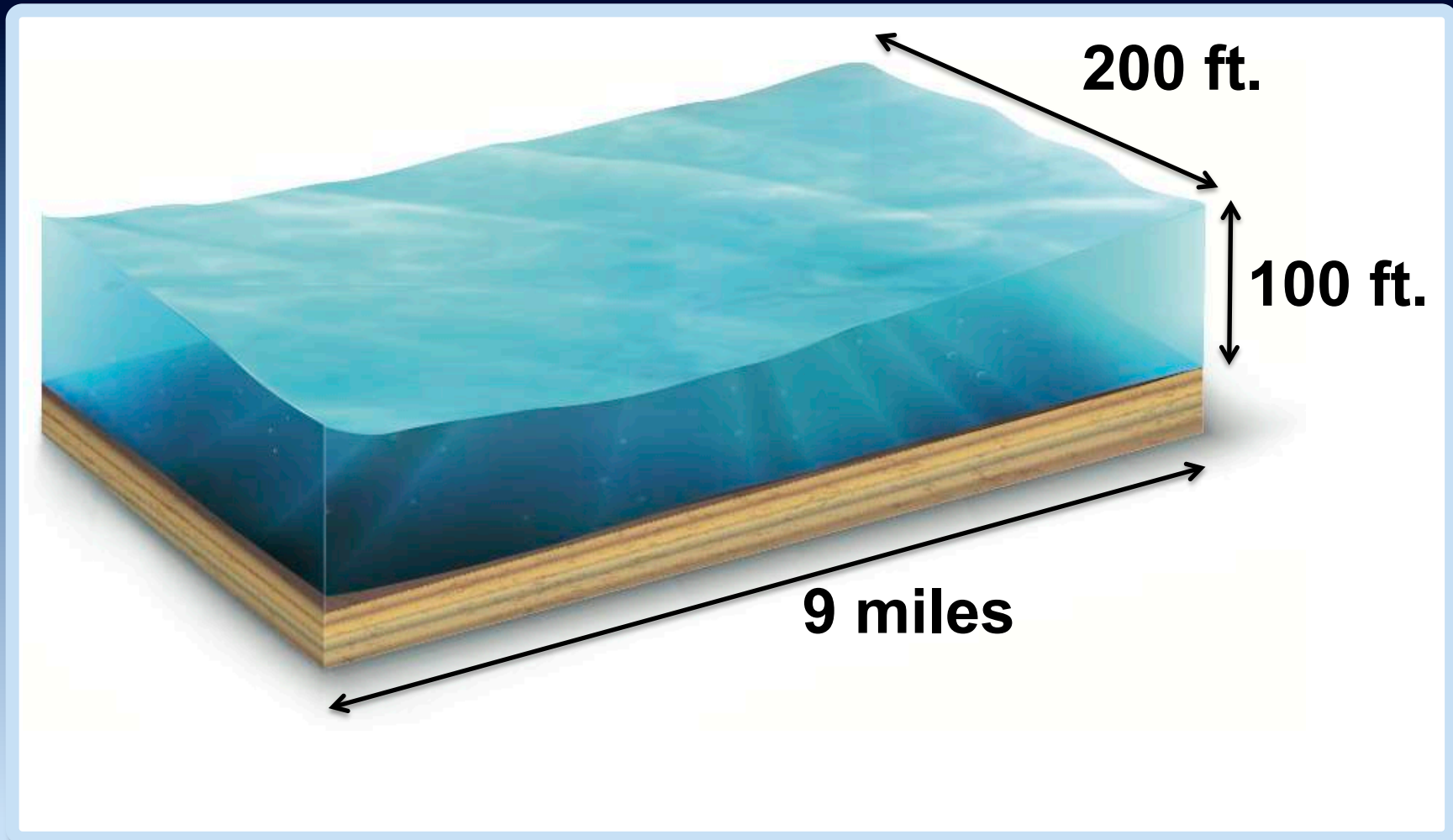
And ...



Going from being a company where most people spent their entire careers using relational databases ... to NoSQL structure, we then ended up creating problems for ourselves ... So with hindsight I would have thought more about the organisational preparedness.

-- Keith Pritchard

Moving corporate data ...



Moving corporate data

- Moving water from one big tank to another without losing a single drop
 - Reading from Relational and writing to NoSQL
- The amount of information currently stored in NoSQL databases would not quench a thirst on a hot day
- Dante has reserved a special place in hell for NoSQL database vendors
 - Moving water from one big tank into another using just a small spoon between their teeth

But ...

- Riak at the National Health Service (UK)
 - New DBMS needs 10-12 people to manage it, compared to over 100 for the old systems
 - Cost of infrastructure supporting new DBMS reduced to ~5% of the old systems
 - Lookup times for patient records significantly reduced from seconds to milliseconds

NoSQL hoopla and hype



Source: Getty Image ID WCO_030



Source: Shutterstock Image ID 92042489



Source: Inspired by "The Next Big Thing 2012" The Wall Street Journal (27 September 2012)



Source: Inspired by “NoSQL takes the database market by storm” Brandon Butler (27 October 2014)



Source: Inspired by <http://www.marketresearchmedia.com/?p=568> and <http://www.pr.com/press-release/613495>



Source: Inspired by <http://dilbert.com/strip/1995-01-22/>





Source: Inspired by <http://vimeo.com/104045795/>



Source: Inspired by <https://www.youtube.com/watch?v=3MNlrKIQp2E>





Source: Inspired by "MongoDB: Second Round" Thomas Jaspers (8 November 2012)



Source: Inspired by “Why MongoDB is Awesome” John Nunemaker (15 May 2010) and “Why Neo4J is awesome in 5 slides” Florent Biville (29 October 2012)





Source: Inspired by <http://slv.io/>



Source: Inspired by "Saturday Night Live" Season 1 Episode 9 (1976)

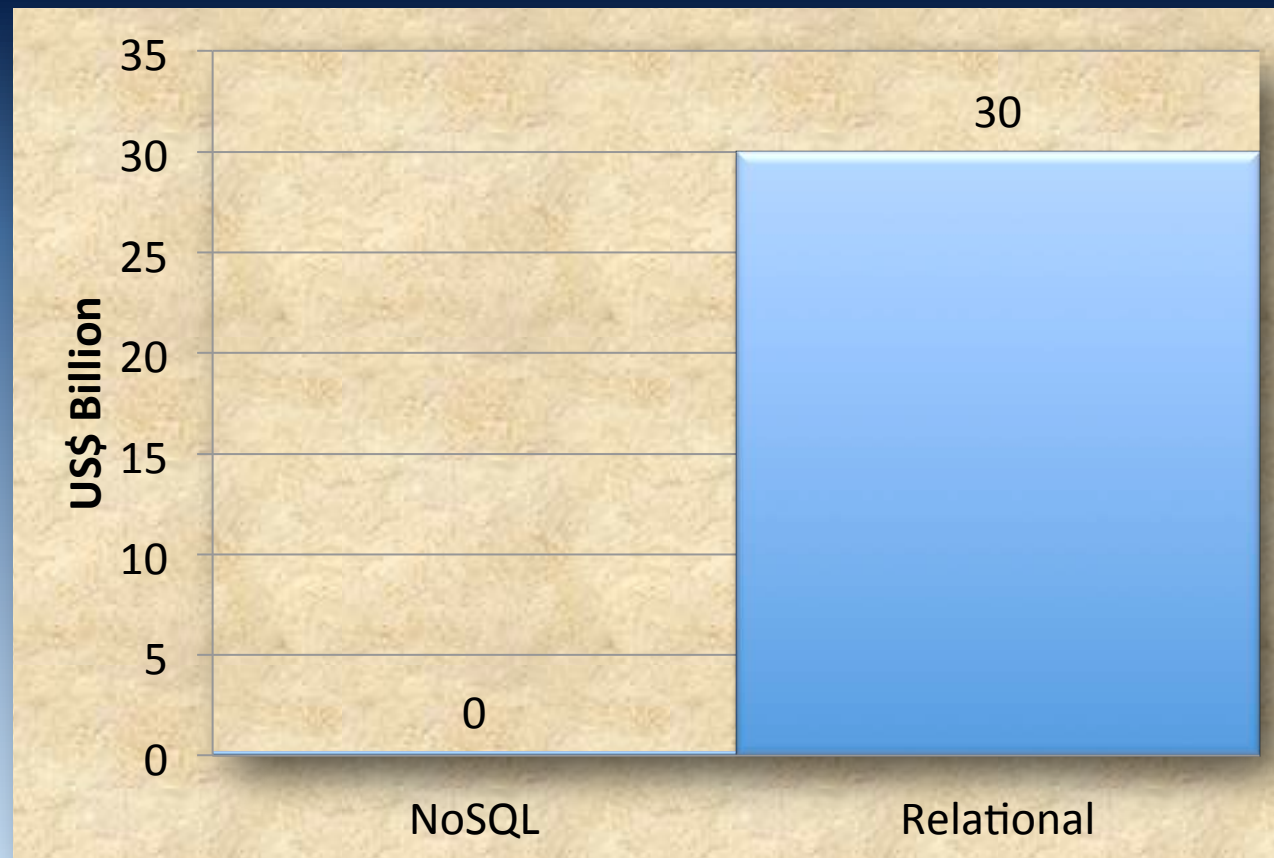
Past proclamations of the imminent demise of relational technology

- Object databases vs. relational
 - GemStone, ObjectStore, Objectivity, etc.
- In-memory databases vs. relational
 - SolidDB, TimesTen, etc.
- Persistence frameworks vs. relational
 - Hibernate, OpenJPA, etc.
- XML databases vs. relational
 - BaseX, Tamino, etc.
- Column-store databases vs. relational
 - Sybase IQ, Vertica, etc.

Market analysis



Database market size ...



Source: "2014 State of Database Technology" InformationWeek (March 2014)

Database market size



NoSQL is a small but growing segment of the database market, according to 451 Research's Matt Aslett, who predicts it at about 2% of the size of the SQL market.

-- Brandon Butler

NoSQL market size

- Private companies do not publish results
- Venture Capital (VC) funding 10s/100s of millions of US \$
- NoSQL revenue
 - \$20 million in 2011^[1]
 - \$184 million in 2012^[2]
 - \$223 million in 2014^[3]

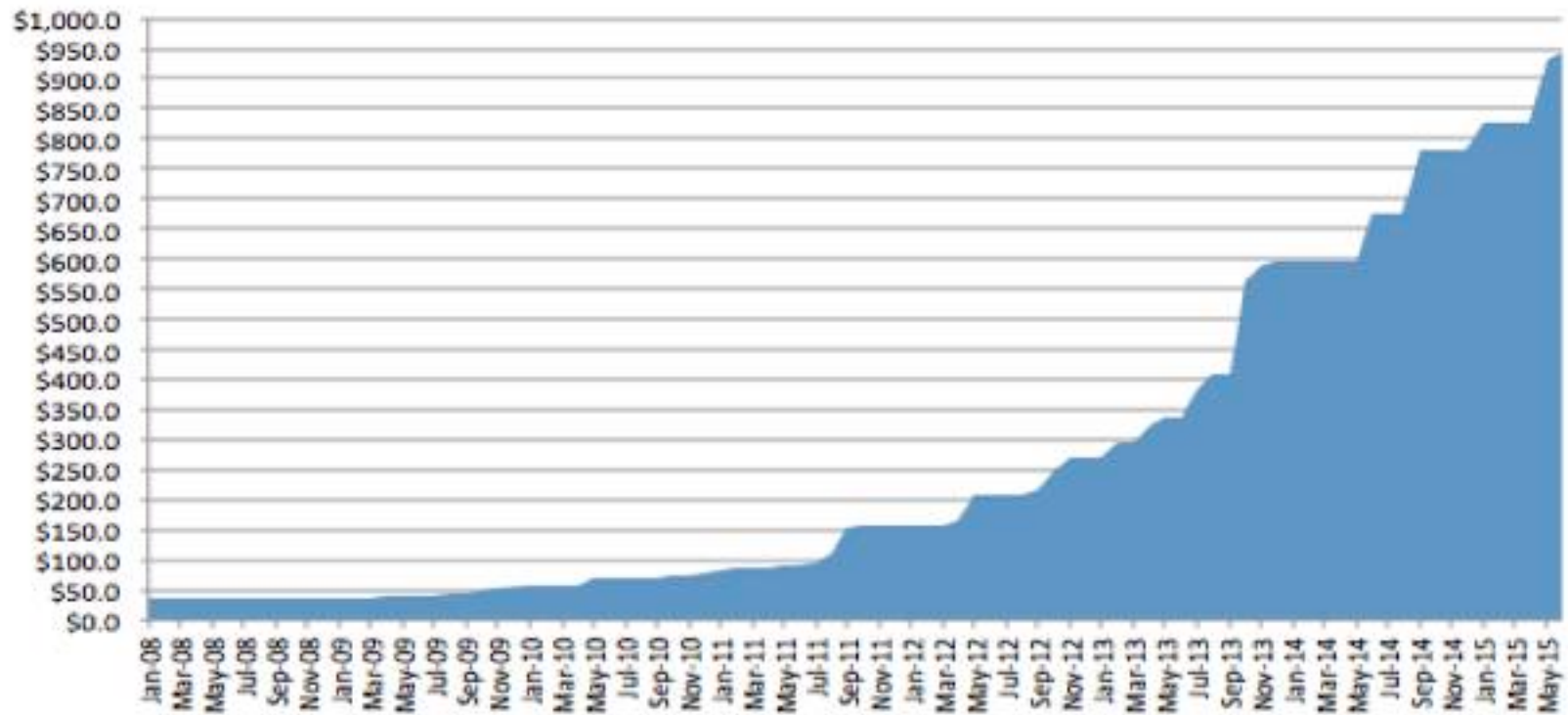


[1] http://blogs.the451group.com/information_management/2012/05/

[2] <http://www.cio.co.uk/insight/data-management/new-database-dawn/>

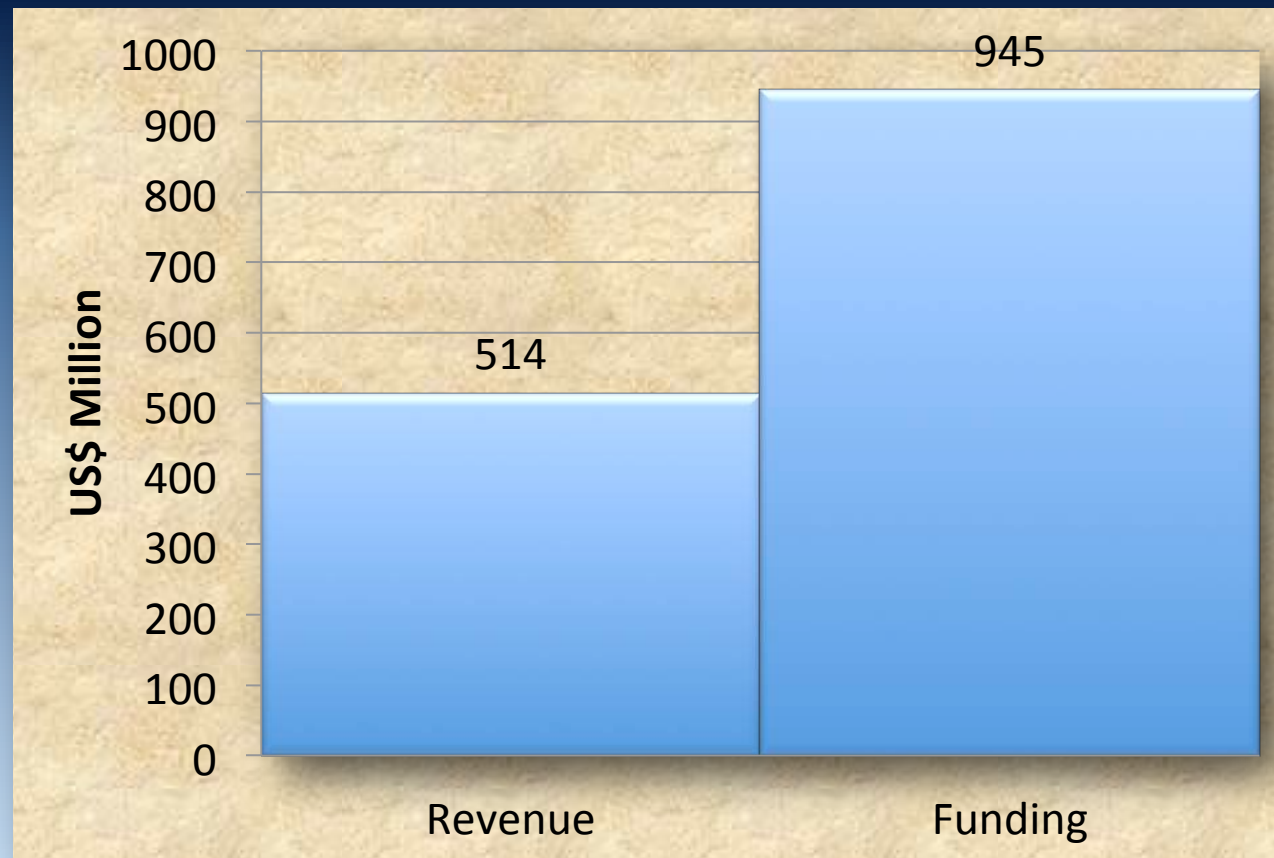
[3] <http://www.datanami.com/2015/04/02/booming-big-data-market-headed-for-60b/>

VC funding for NoSQL vendors, Jan 2008-Jun 2015



Source: "NoSQL by the numbers" Matt Aslett (23 July 2015)

2014 revenue vs. funding



Source: "NoSQL by the numbers" Matt Aslett (23 July 2015)

Investment in NoSQL, NewSQL

Company	\$ (Million)
MongoDB	231
Couchbase	116
DataStax	83.7
Clustrix	59.3
Basho	32.5
FoundationDB	22.3
Aerospike	22

Source: "The NoSQLNow conference in San Jose this week" Jnan Dash (22 August 2014)

Recent investment in NoSQL

Company	\$ (Million)
MongoDB	311 ^[1]
DataStax	189.7 ^[1]
MarkLogic	173 ^[2]
Couchbase	116
Basho	64 ^[3]
Neo4j	44.1 ^[4]
Redis Labs	28 ^[5]

[1] <http://venturebeat.com/2015/01/12/basho-funding/>

[2] <http://fortune.com/2015/05/12/marklogic-snags-102-million/>

[3] <http://www.idgconnect.com/abstract/9332/basho-enterprise-focus-winning-friends-funds/>

[4] <http://fortune.com/2015/02/03/datastax-acquisition-database-software/>

[5] <http://www.informationweek.com/big-data/big-data-analytics/redis-emerges-as-nosql-in-memory-performer-/d/d-id/1321047>

Vendor revenue example ...



The new funding, which values MongoDB at \$1.6 billion ... Wikibon estimates MongoDB's 2014 revenue at \$46 million, meaning the company is valued at approximately 35-times lagging 12-month revenue ...

-- Jeff Kelly

Vendor revenue example



MongoDB ... I would say if we could get to 20 to 25 per cent of our user base then we would have a multi-billion dollar company; [at the moment] it's less than five per cent
-- Dev Ittycheria

Vendor profitability example



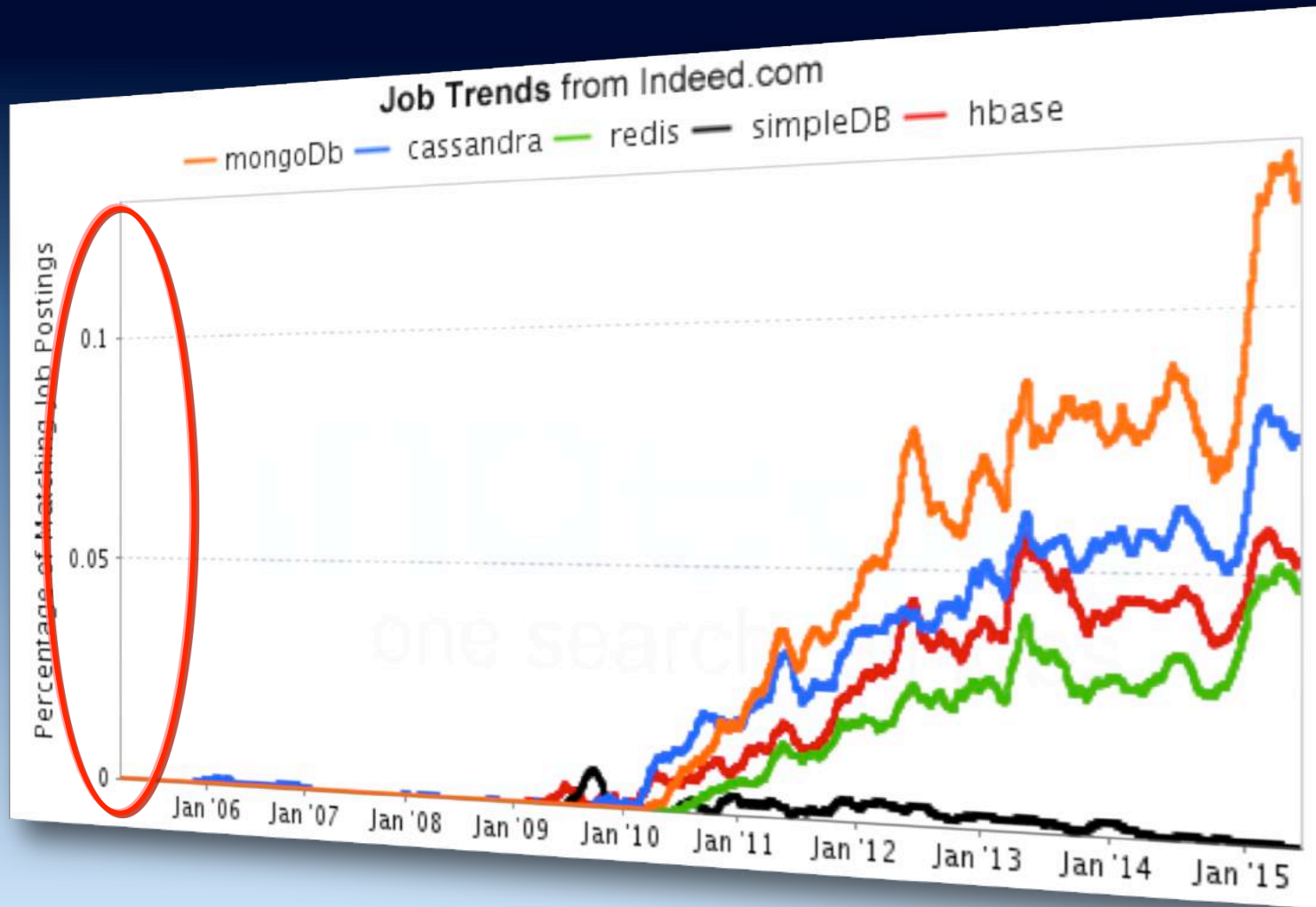
MongoDB ... Profitability is still at least a couple years away, Chairman and Co-founder Dwight Merriman told me in an interview.

-- Ben Fischer

Number of customers

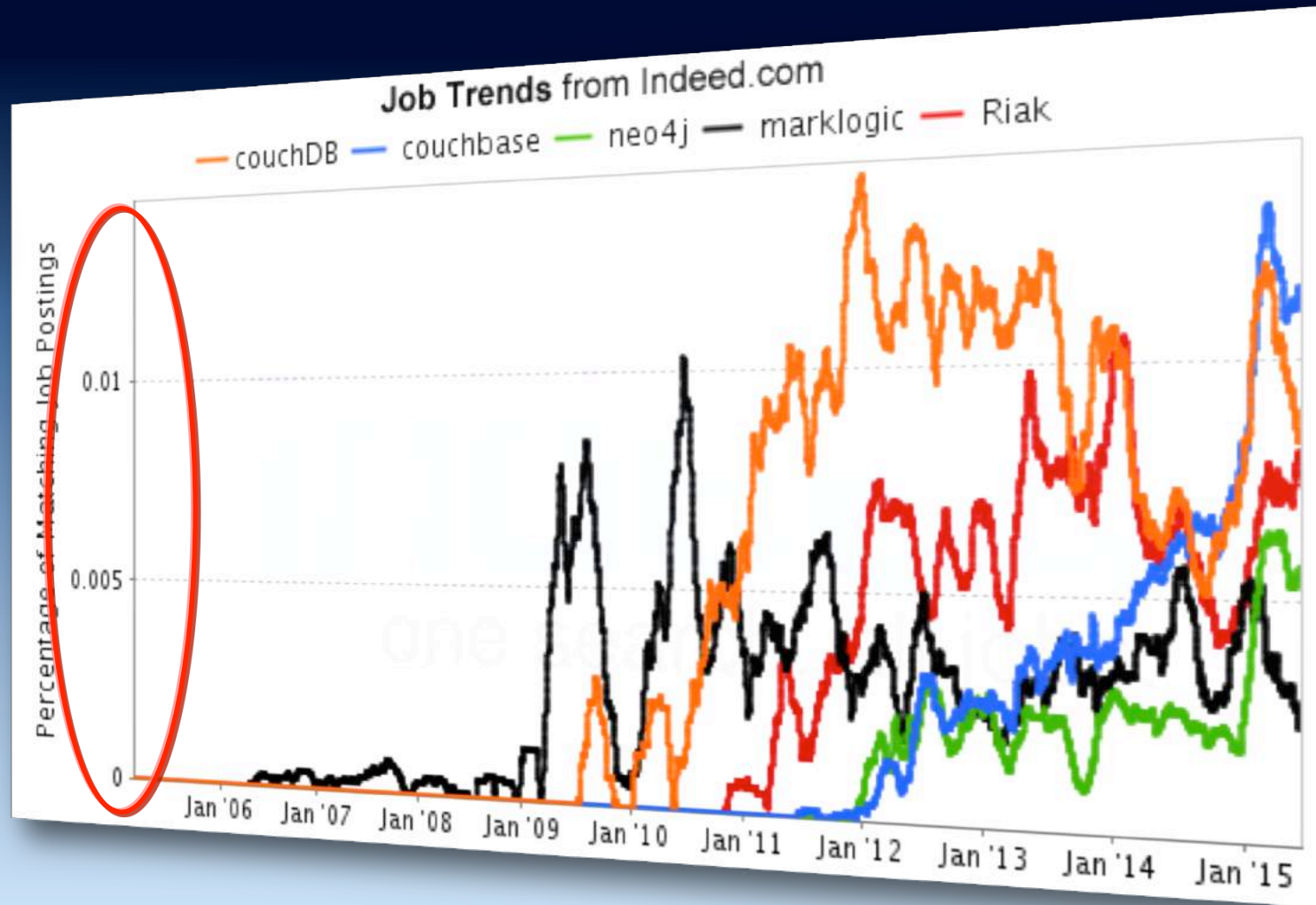
Company	Customers
MongoDB	2500
DataStax	500
MarkLogic	500
Couchbase	450
Basho	200
Neo4j	150
Total	4300

NoSQL job trends ...



Source: After “NoSQL Job Trends: August 2014” Robert Diana (4 September 2014)

NoSQL job trends ...



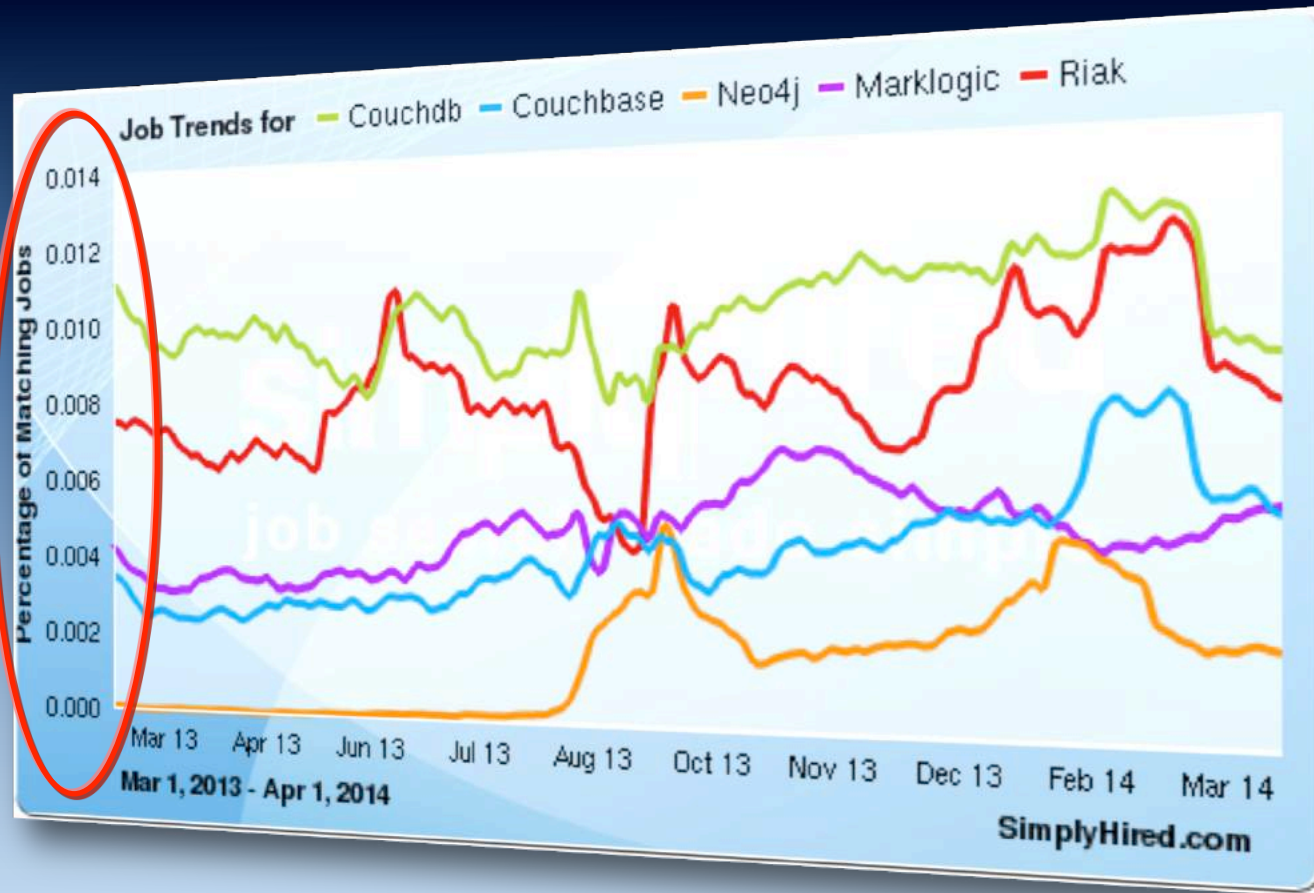
Source: After “NoSQL Job Trends: August 2014” Robert Diana (4 September 2014)

NoSQL job trends ...



Source: "NoSQL Job Trends: August 2014" Robert Diana (4 September 2014)

NoSQL job trends



Source: "NoSQL Job Trends: August 2014" Robert Diana (4 September 2014)

Most valuable IT skills in 2012

Skill	\$
1. Hadoop	115,062
2. Big Data	113,739
3. NoSQL	113,031
4. PMBook	110,885
5. Omnigraffle	110,758
6. SOA	109,504
7. Mongo DB	108,304
8. Jetty	106,936
9. Objective C	104,989
10. ETL	104,777

Source: "Dice Tech Salary Survey" Dice (22 January 2013)

Most valuable IT skills in 2013

Skill	\$
1. R	115,531
2. NoSQL	114,796
3. MapReduce	114,396
4. PMBook	112,382
5. Cassandra	112,382
6. Omnigraffle	111,039
7. Pig	109,561
8. SOA	108,997
9. Hadoop	108,669
10. Mongo DB	107,825

Source: "Dice Tech Salary Survey" Dice (29 January 2014)

Most valuable IT skills in 2014

Skill	\$
1. PaaS	130,081
2. Cassandra	128,646
3. MapReduce	127,315
4. Cloudera	126,816
5. HBase	126,369
6. Pig	124,563
7. ABAP	124,262
8. Chef	123,458
9. Flume	123,186
10. Hadoop	121,313

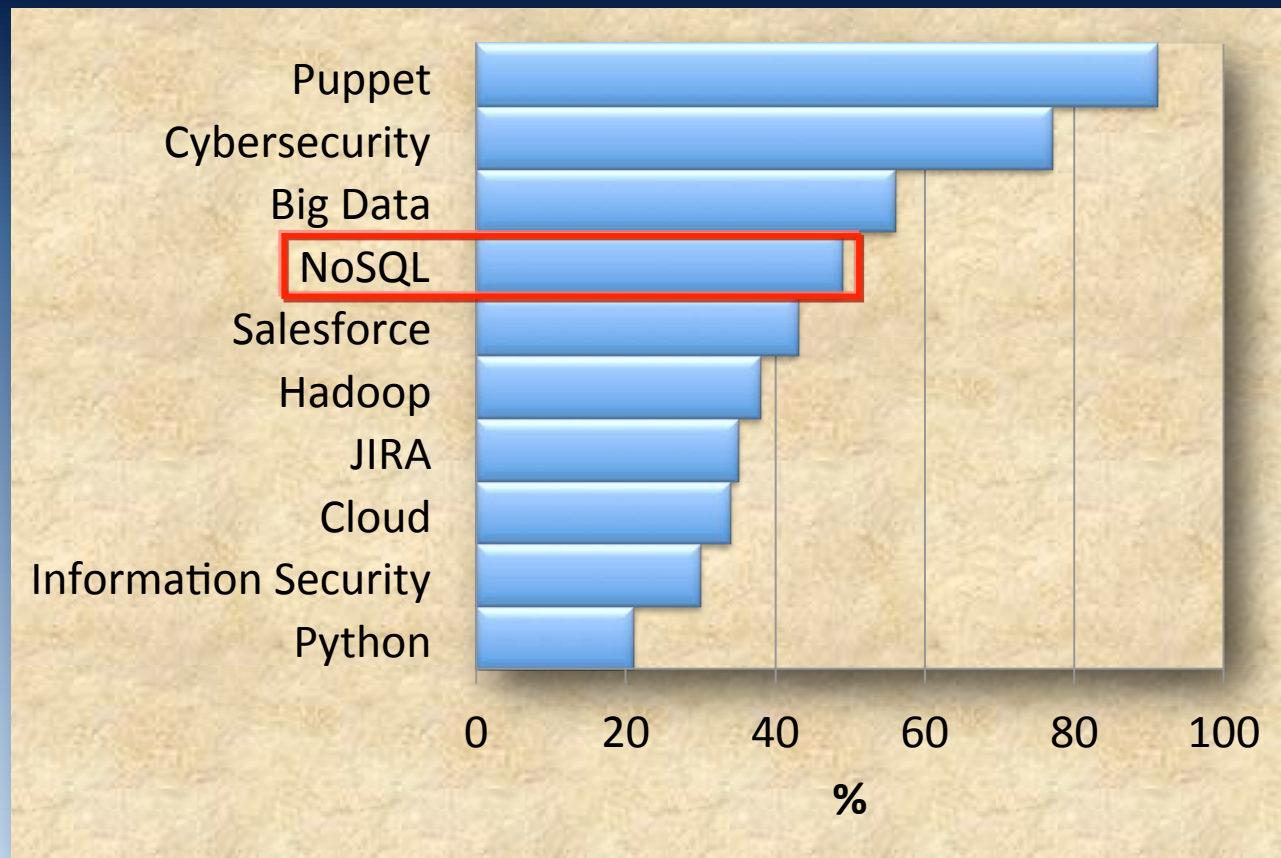
Source: "Dice Tech Salary Survey" Dice (22 January 2015)

Most valuable IT skills in 2015

Skill	\$
1. HANA	154,749
2. Cassandra	147,811
3. Cloudera	142,835
4. PaaS	140,894
5. OpenStack	138,579
6. CloudStack	138,095
7. Chef	136,850
8. Pig	132,850
9. MapReduce	131,563
10. Puppet	131,121

Source: "Dice Tech Salary Survey" Dice (26 January 2016)

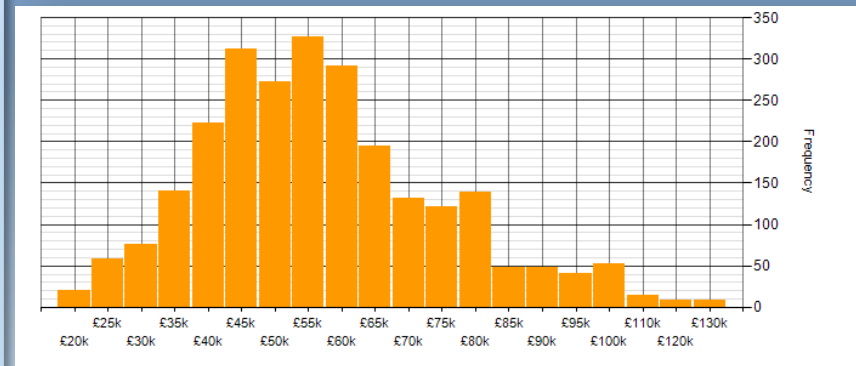
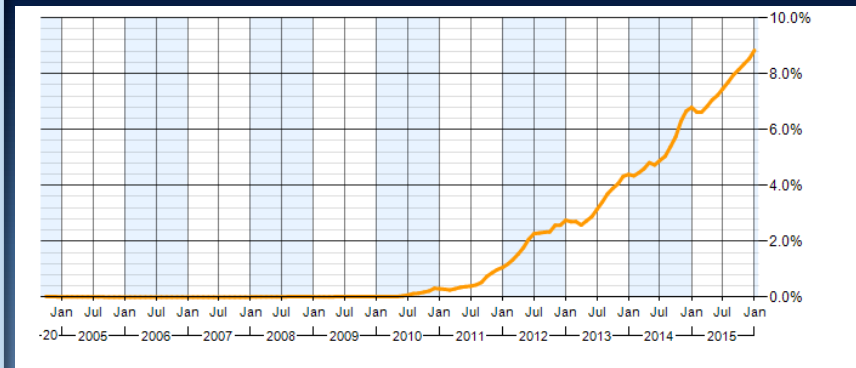
Fastest growing tech skills



Source: "The Fastest-Growing Tech Skills: Dice Report" Shravan Goli (15 September 2014)

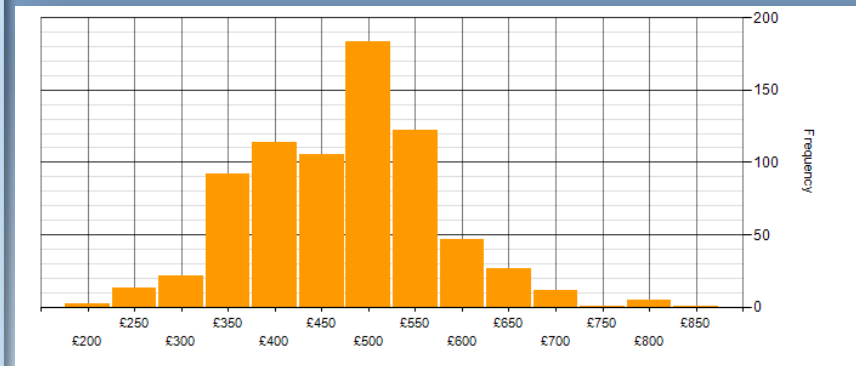
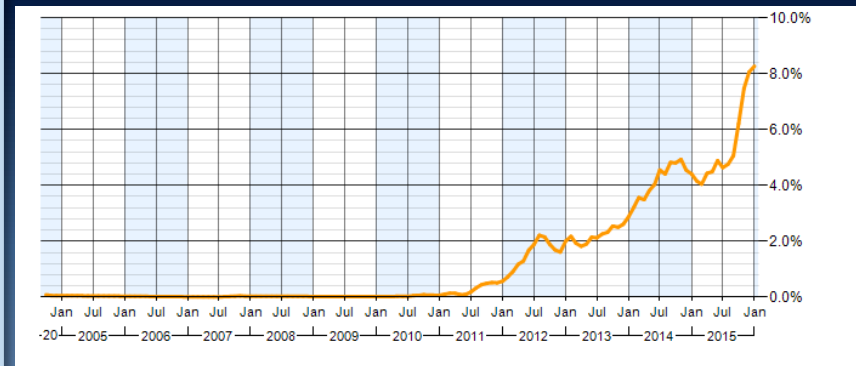
NoSQL jobs in the UK (perm)

- Database and Business Intelligence
 - MongoDB (1892)
 - Cassandra (871)
 - Redis (338)
 - Neo4j (183)
 - CouchDB (181)
 - Couchbase (174)
 - HBase (158)
 - Riak (144)

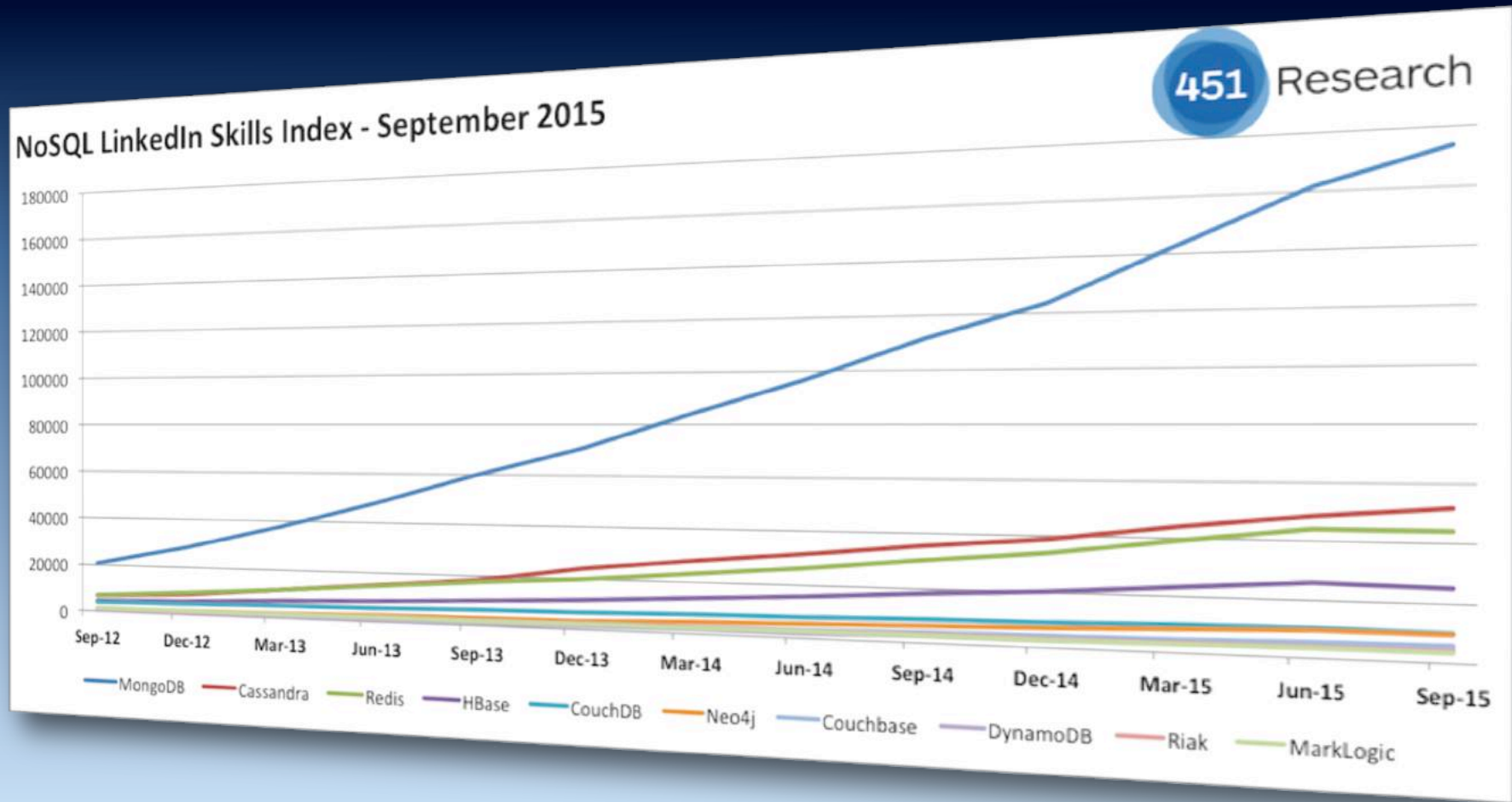


NoSQL jobs in the UK (contract)

- Database and Business Intelligence
 - MongoDB (746)
 - Cassandra (392)
 - Redis (133)
 - HBase (65)
 - CouchDB (55)
 - DynamoDB (52)
 - Couchbase (43)
 - Neo4j (31)

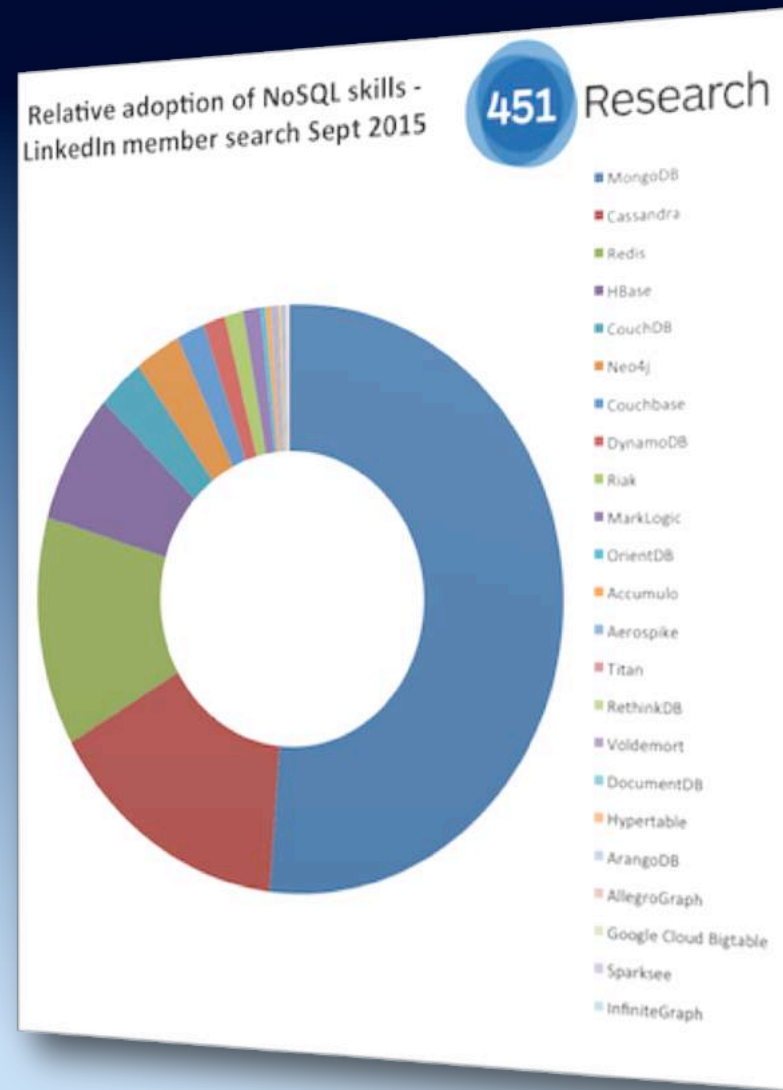


NoSQL LinkedIn skills index ...



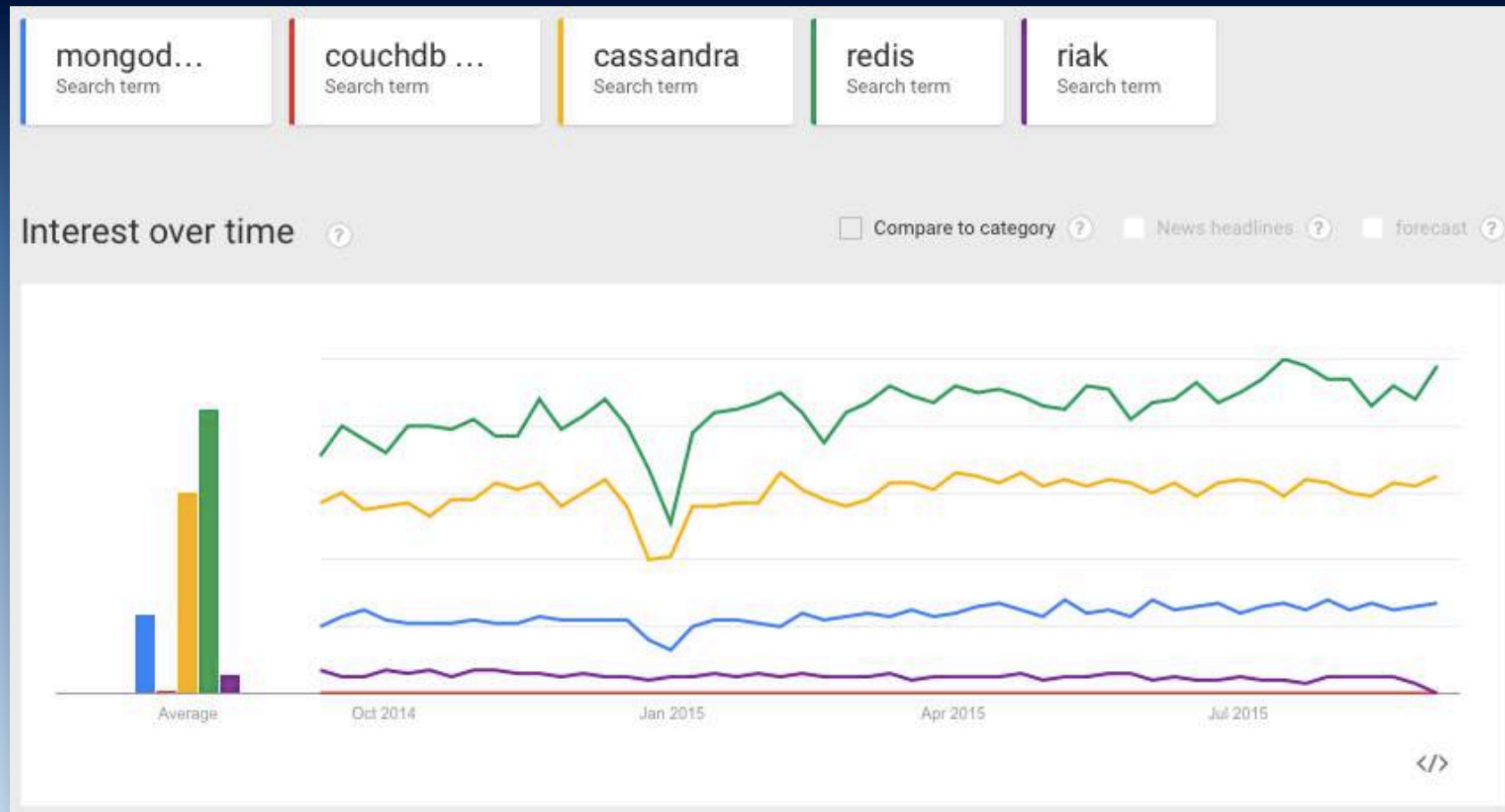
Source: "NoSQL LinkedIn Skills Index - September 2015" Matthew Aslett (1 October 2015)

NoSQL LinkedIn skills index



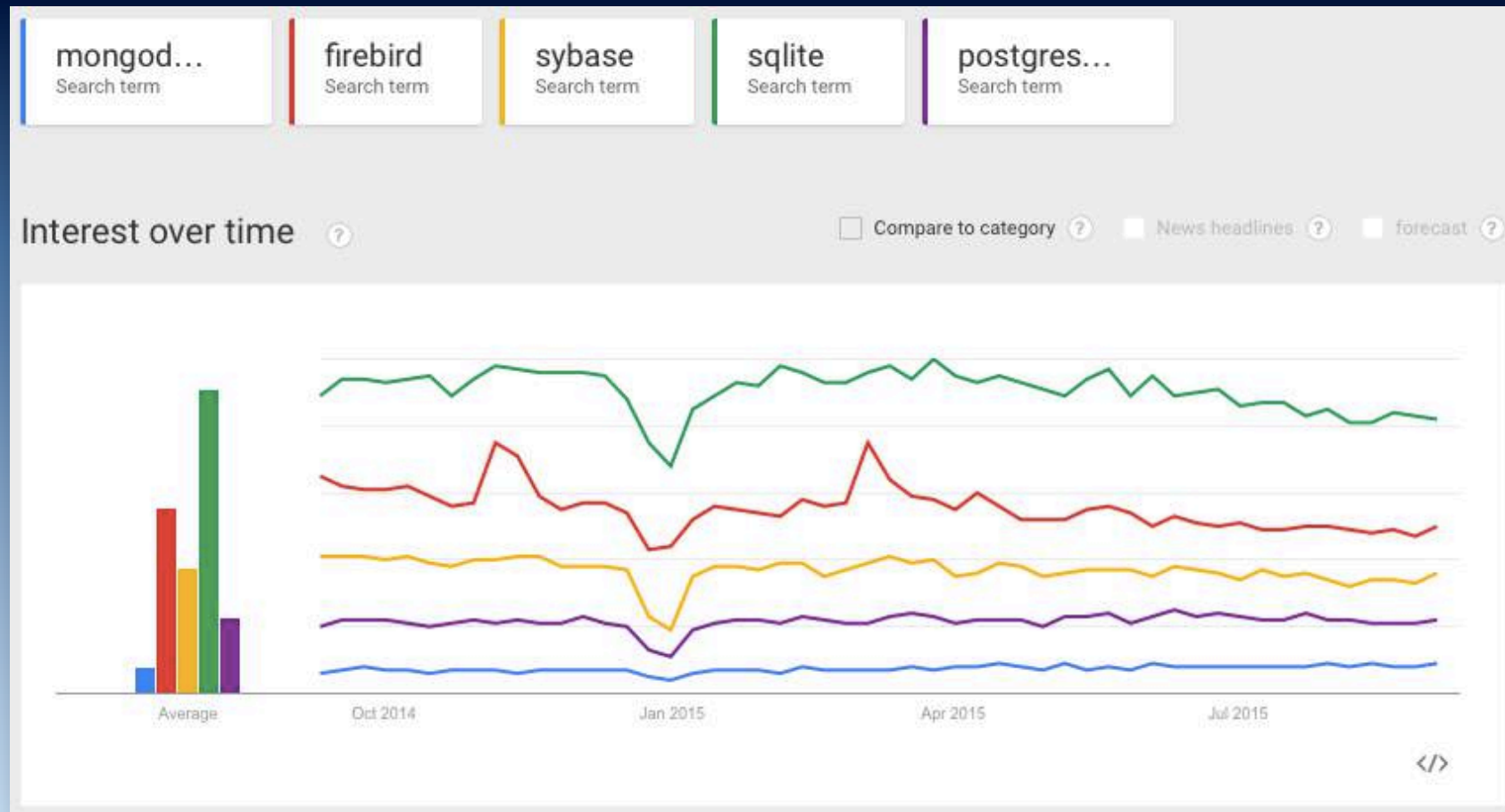
Source: "NoSQL LinkedIn Skills Index - September 2015" Matthew Aslett (1 October 2015)

NoSQL vs. the world ...



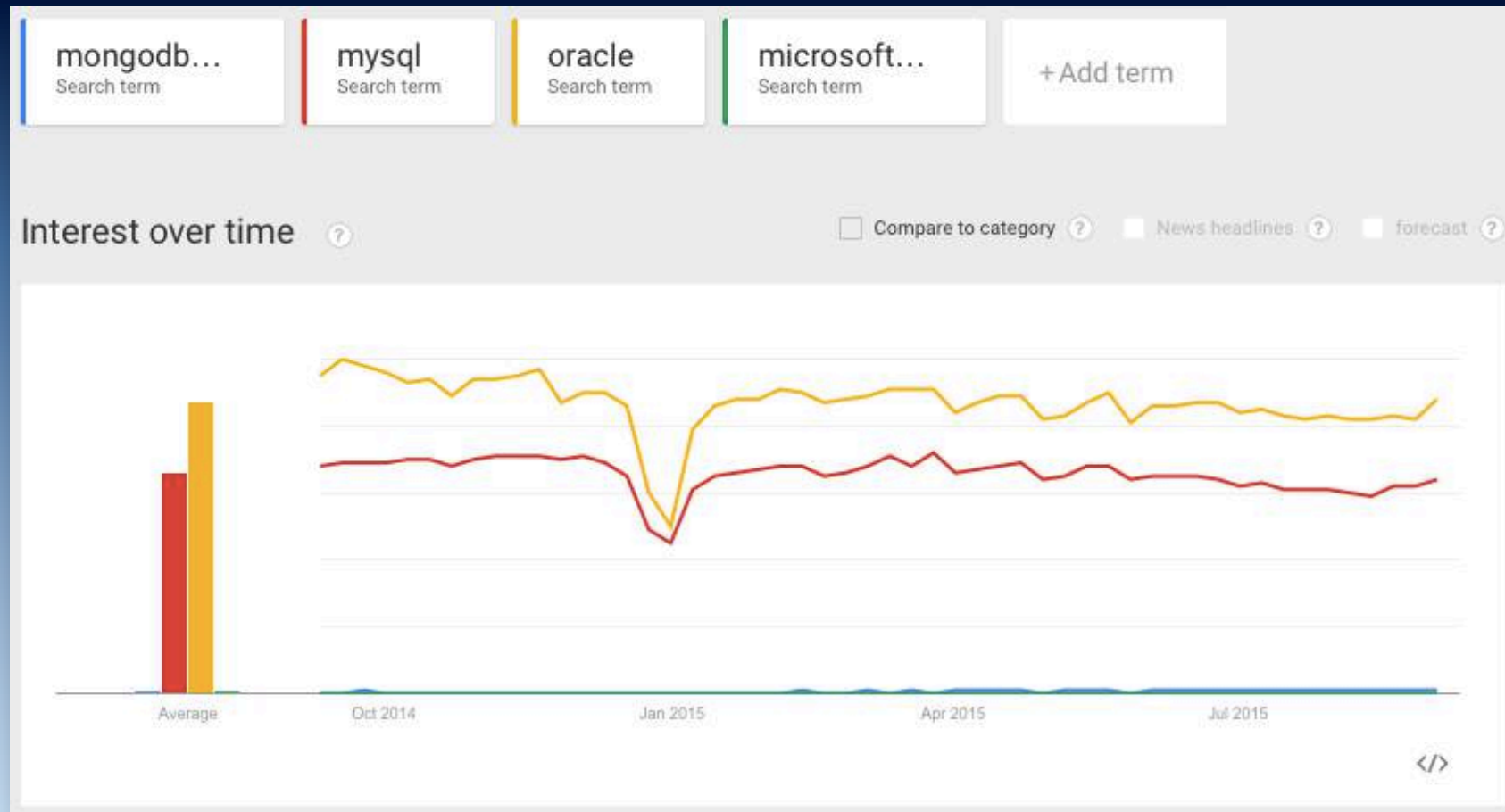
Source: After "NoSQL vs. the world" Kristina Chodorow (5 May 2011)

NoSQL vs. the world ...



Source: After "NoSQL vs. the world" Kristina Chodorow (5 May 2011)

NoSQL vs. the world



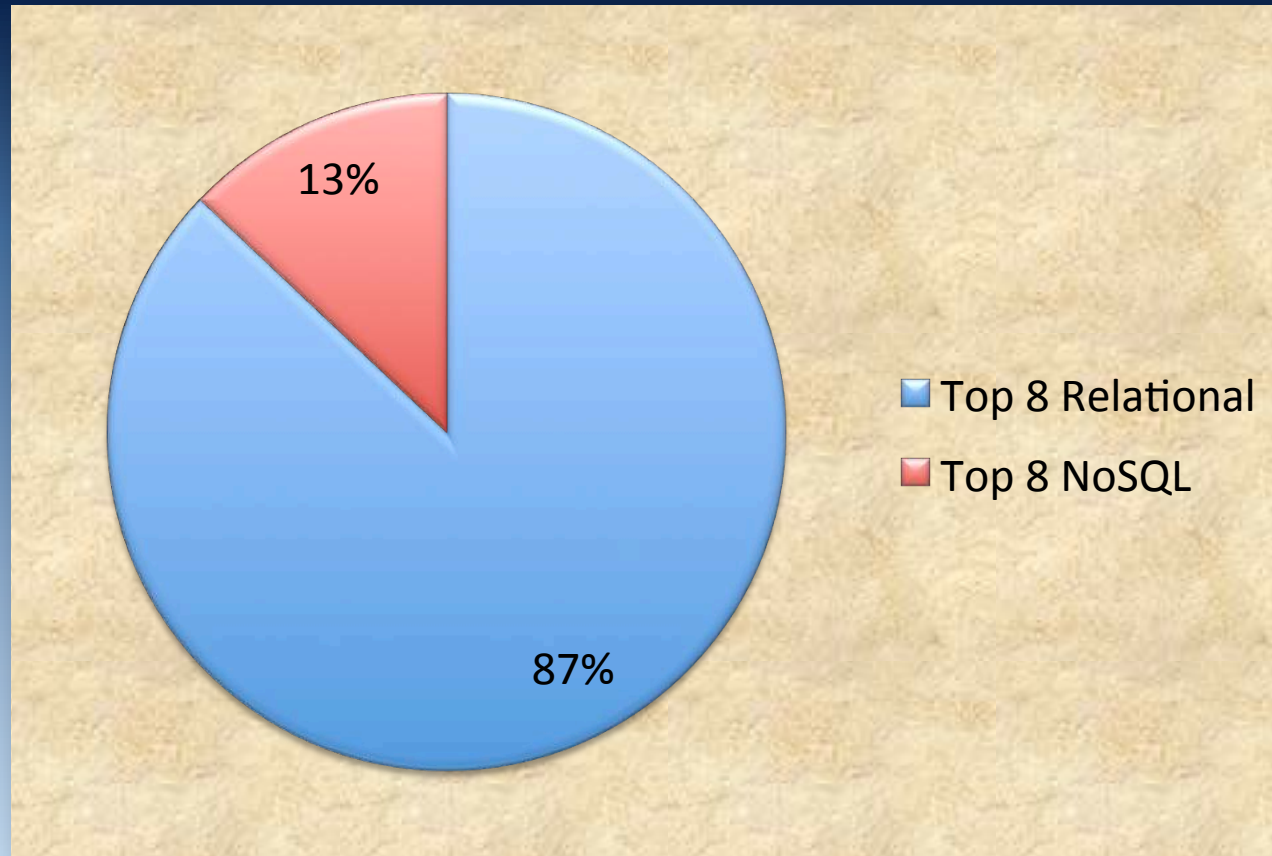
Source: After “NoSQL vs. the world” Kristina Chodorow (5 May 2011)

DB-Engines ranking ...



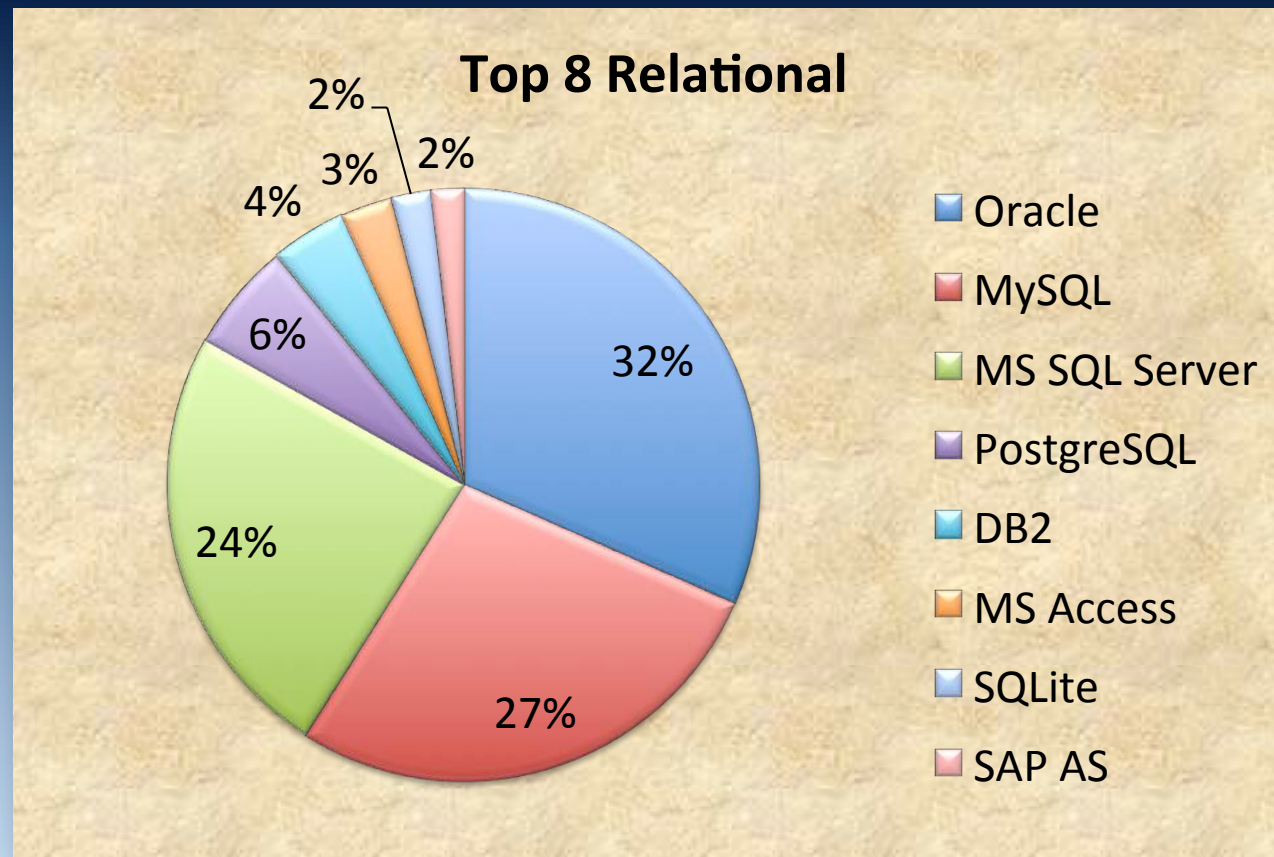
Source: http://db-engines.com/en/ranking_trend/ (4 September 2015)

DB-Engines ranking ...



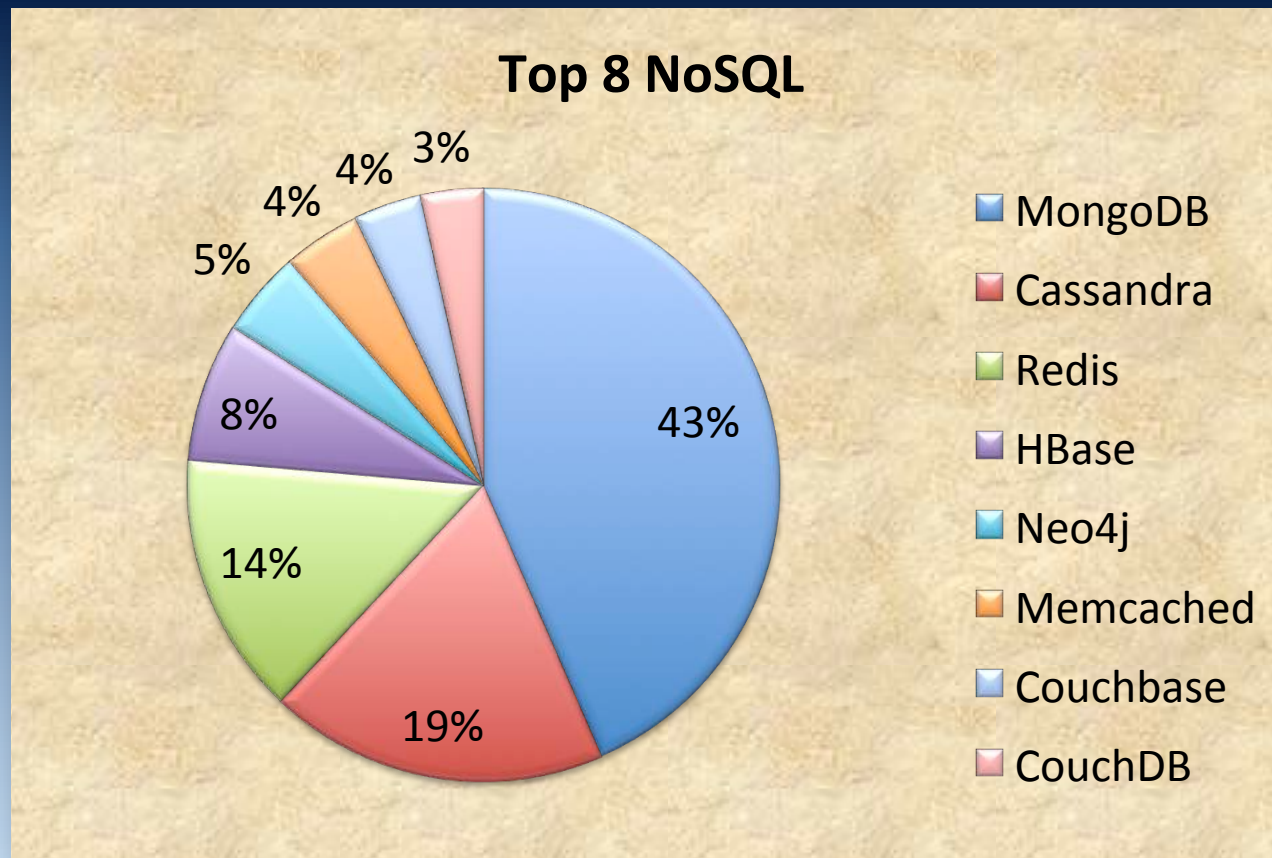
Source: <http://db-engines.com/en/ranking/> (30 January 2016)

DB-Engines ranking ...



Source: <http://db-engines.com/en/ranking/> (30 January 2016)

DB-Engines ranking



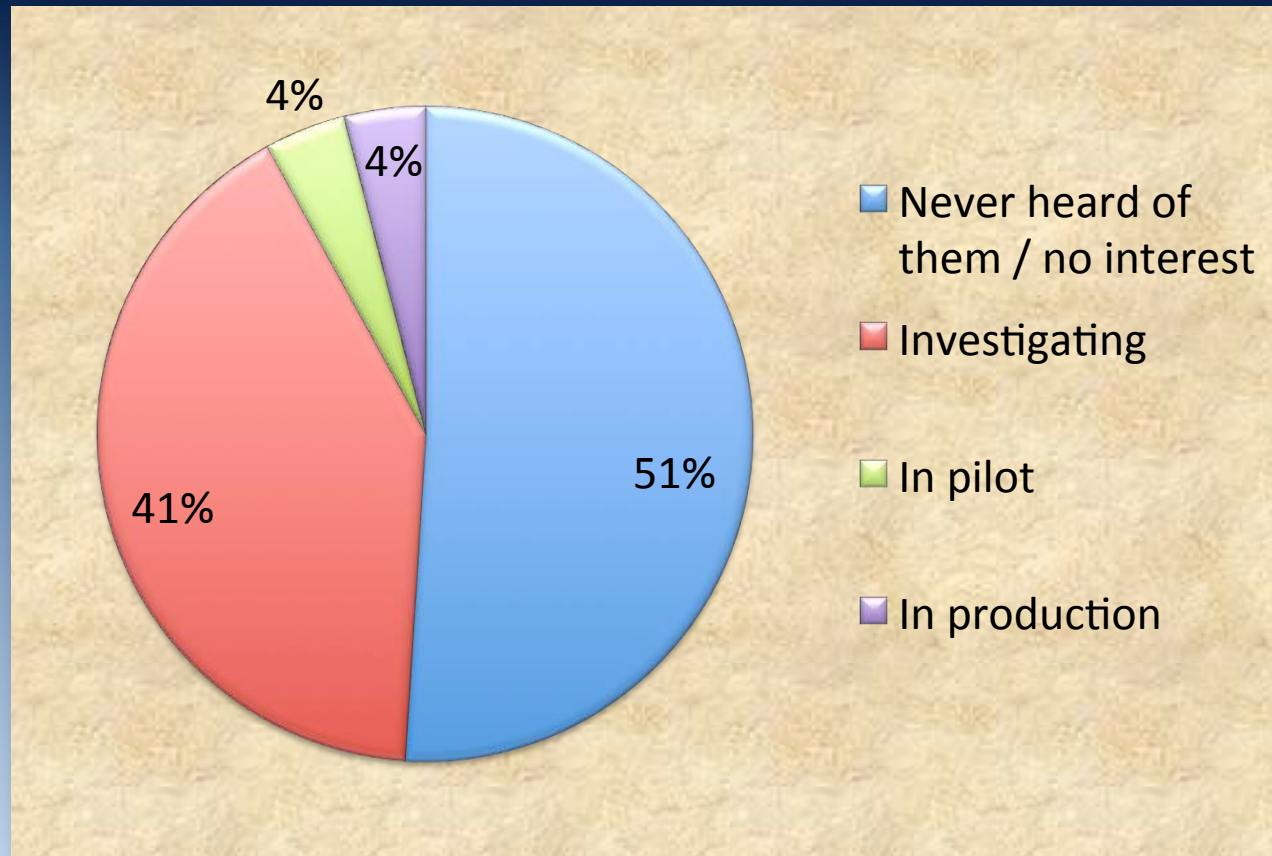
Source: <http://db-engines.com/en/ranking/> (30 January 2016)

But ...



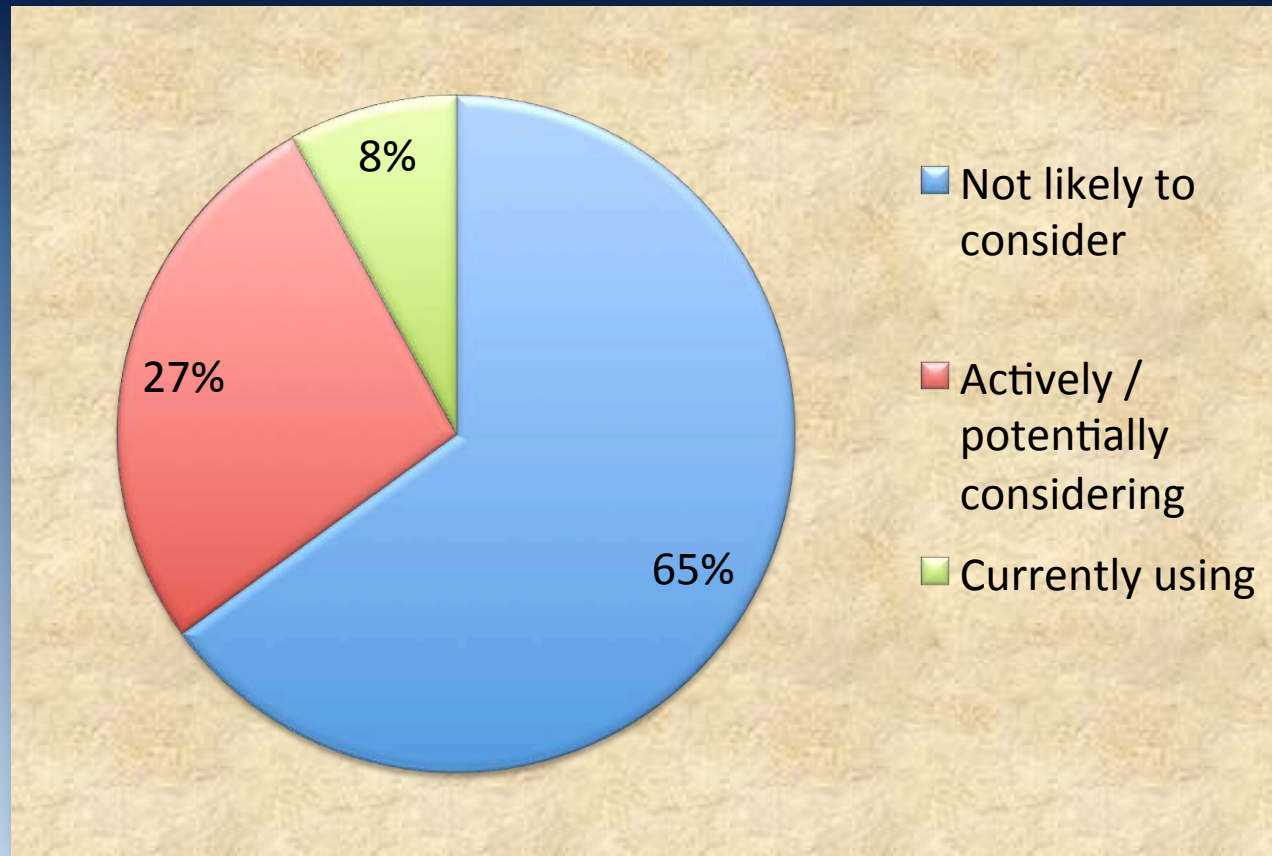
DB-Engines.com ... a popularity rating based on web mentions/searches and installation numbers are not the same thing ...

Use of NoSQL products



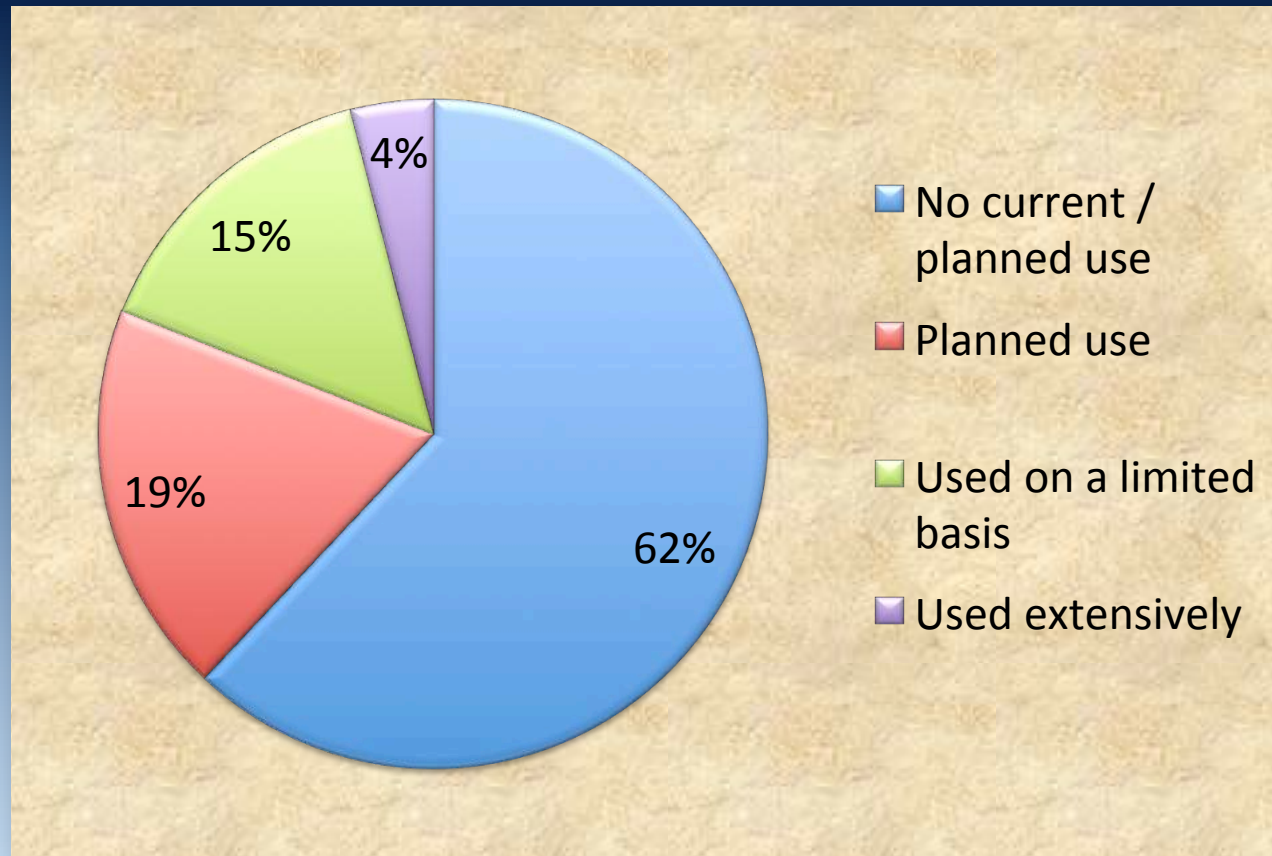
Source: "State of Database Technology 2013" InformationWeek (April 2013)

NoSQL in enterprise apps



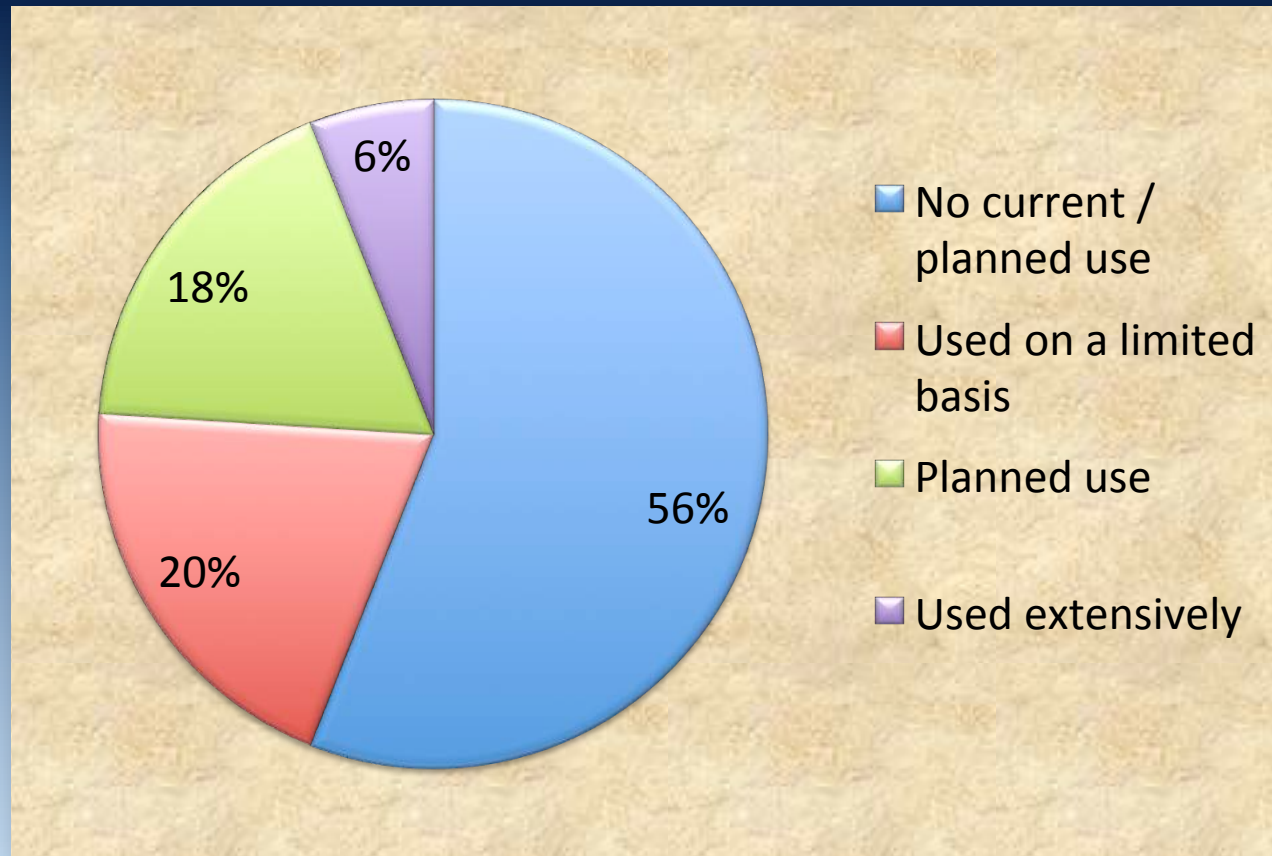
Source: "Cloud Software: Where Next?" InformationWeek (August 2013)

NoSQL in use 2013



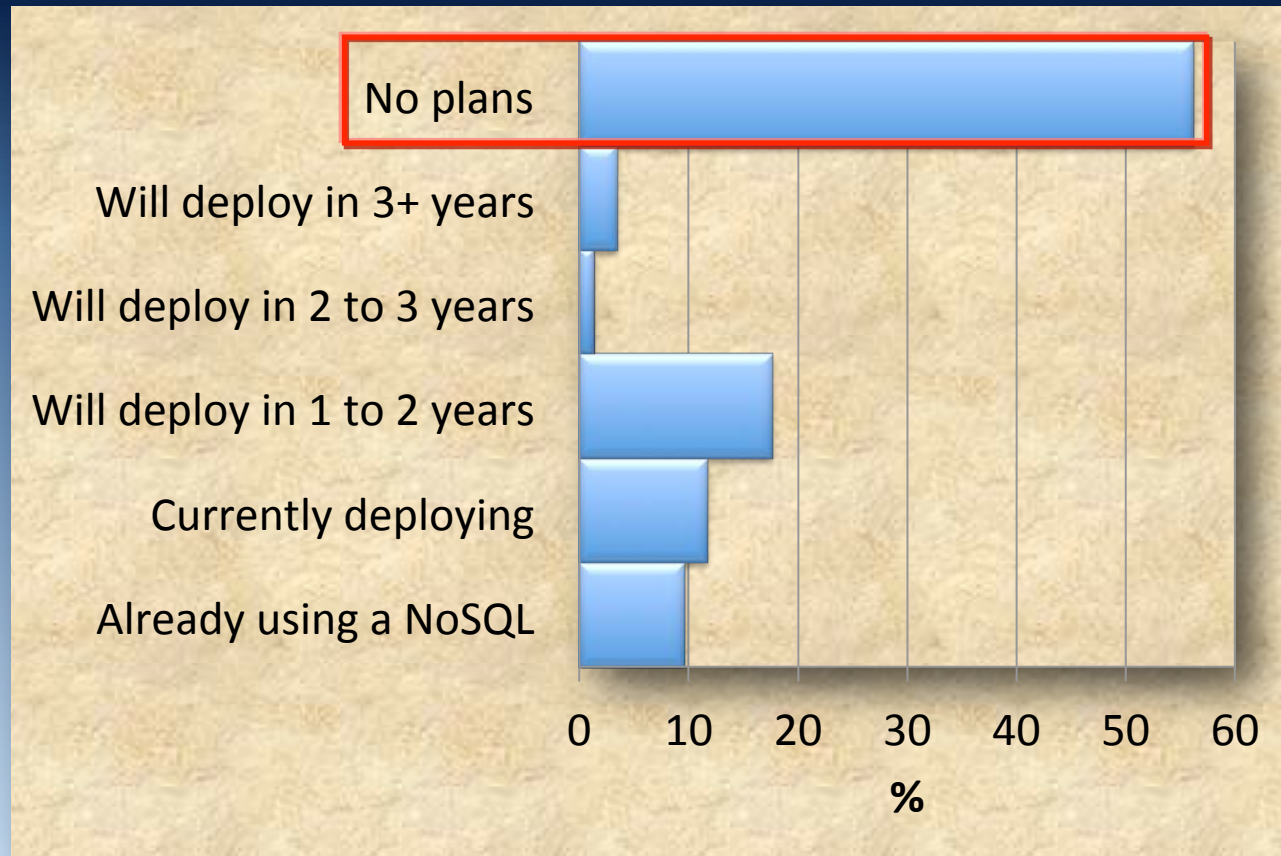
Source: "2014 Analytics, BI, and Information Management Survey" InformationWeek (November 2013)

NoSQL in use 2014



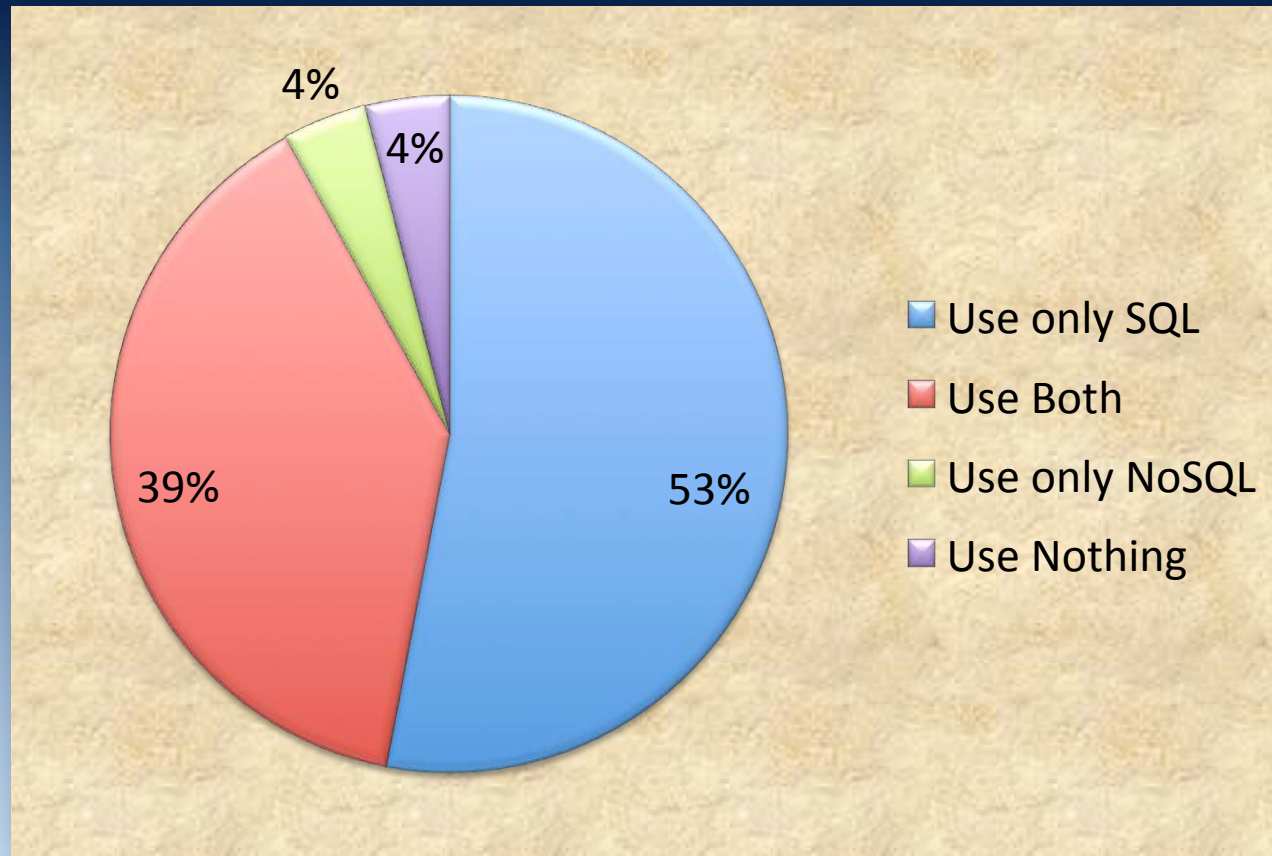
Source: "2015 Analytics & BI Survey" InformationWeek (December 2014)

Does your company currently have plans to adopt NoSQL?



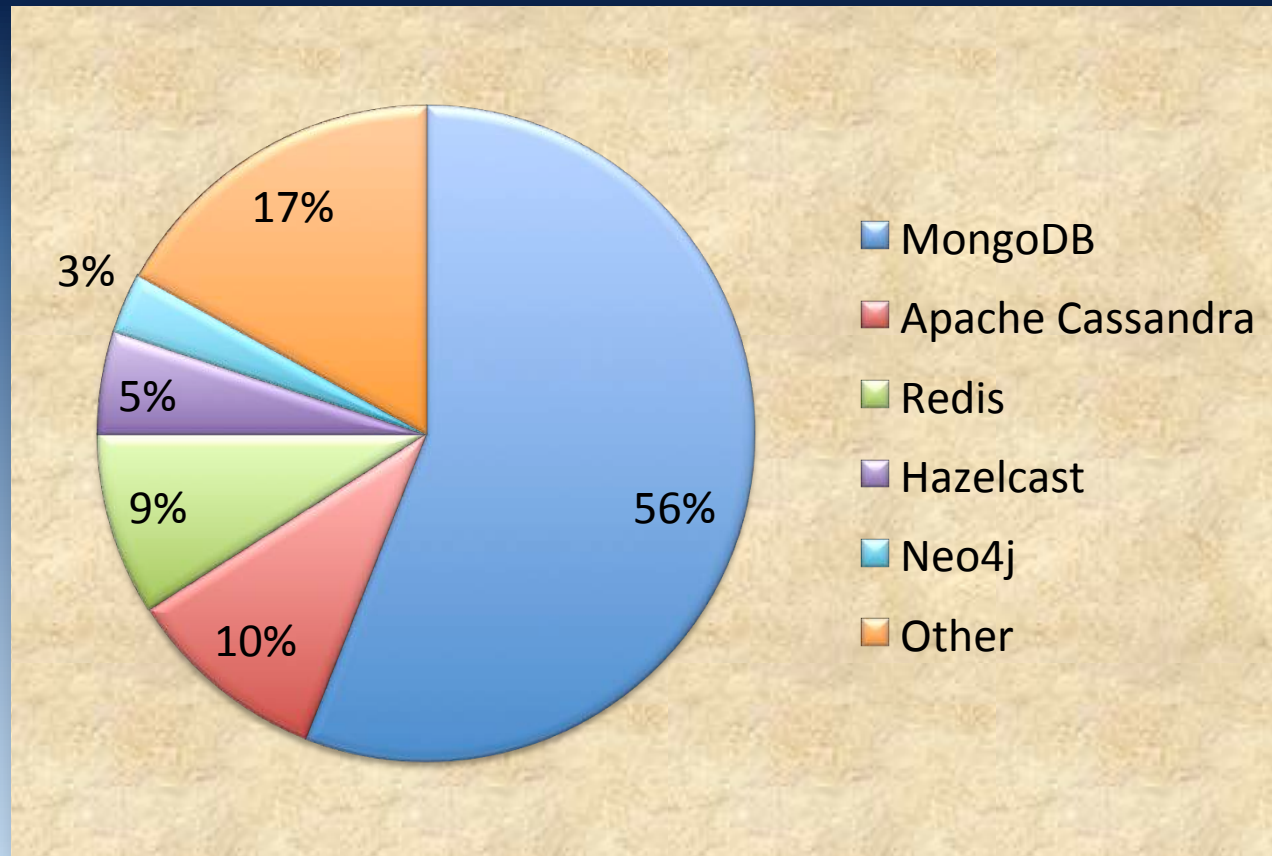
Source: "The Real World of The Database Administrator" Elliot King (March 2015)

SQL, NoSQL or both?



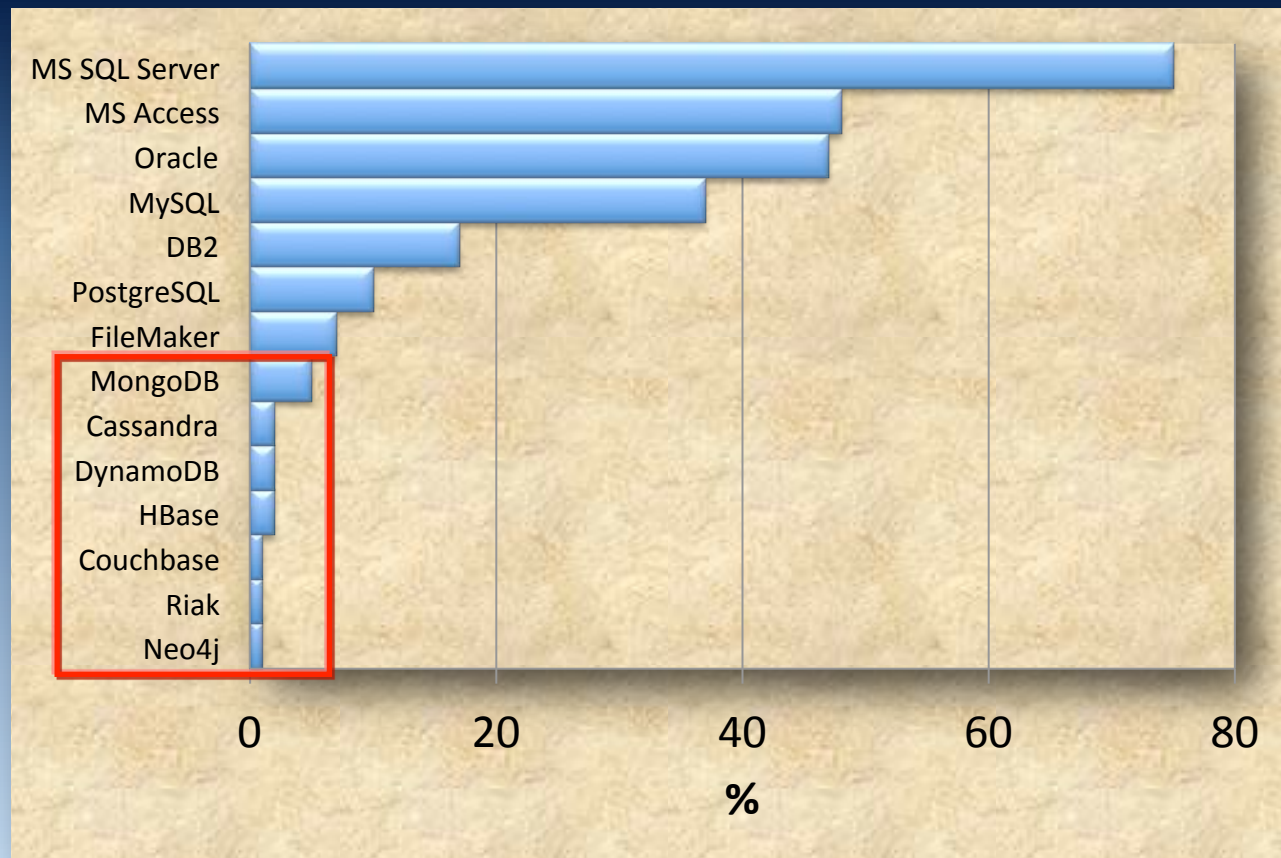
Source: "Java Tools & Technologies Landscape for 2014" ZeroTurnaround (May 2014)

Primary NoSQL technology



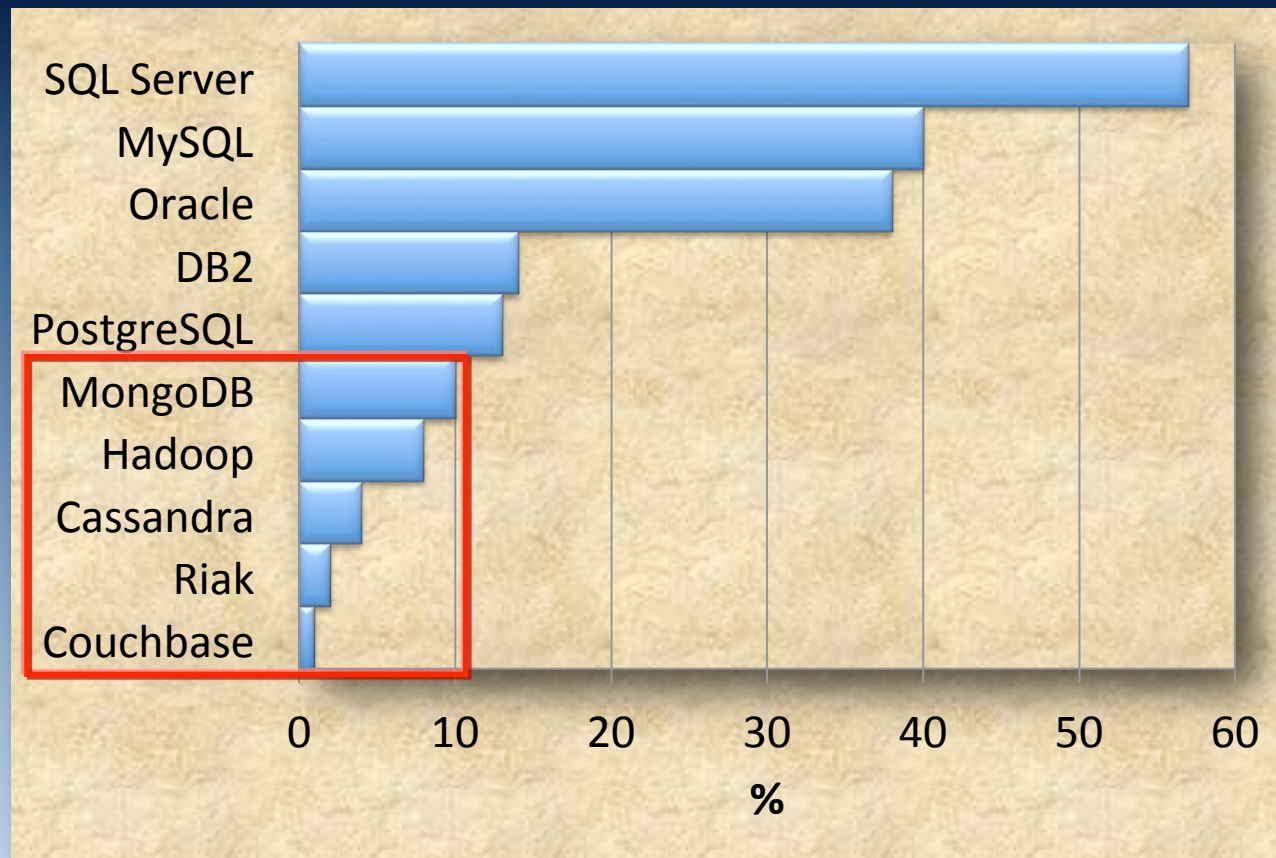
Source: "Java Tools & Technologies Landscape for 2014" ZeroTurnaround (May 2014)

Databases in use

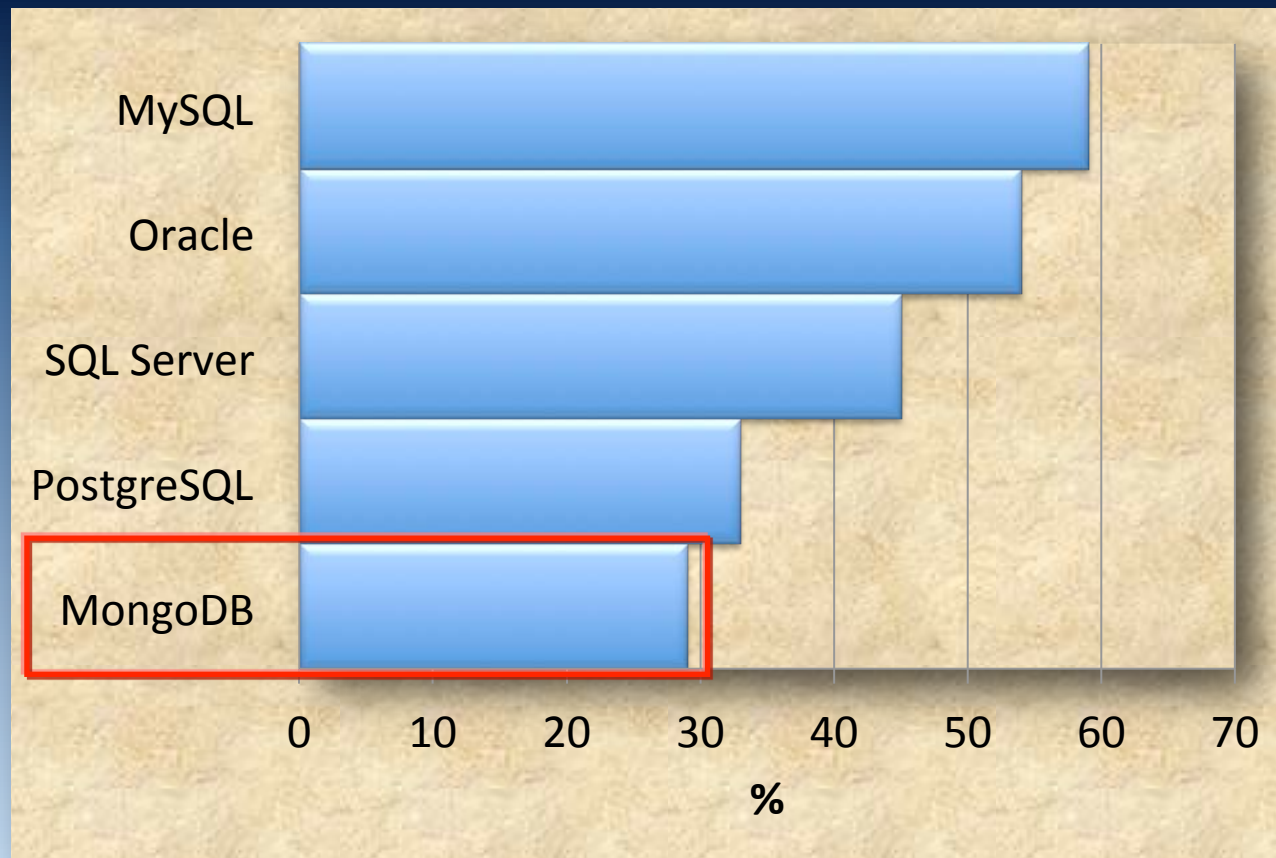


Source: "2014 State of Database Technology" InformationWeek (March 2014)

What database(s) does your company currently use?

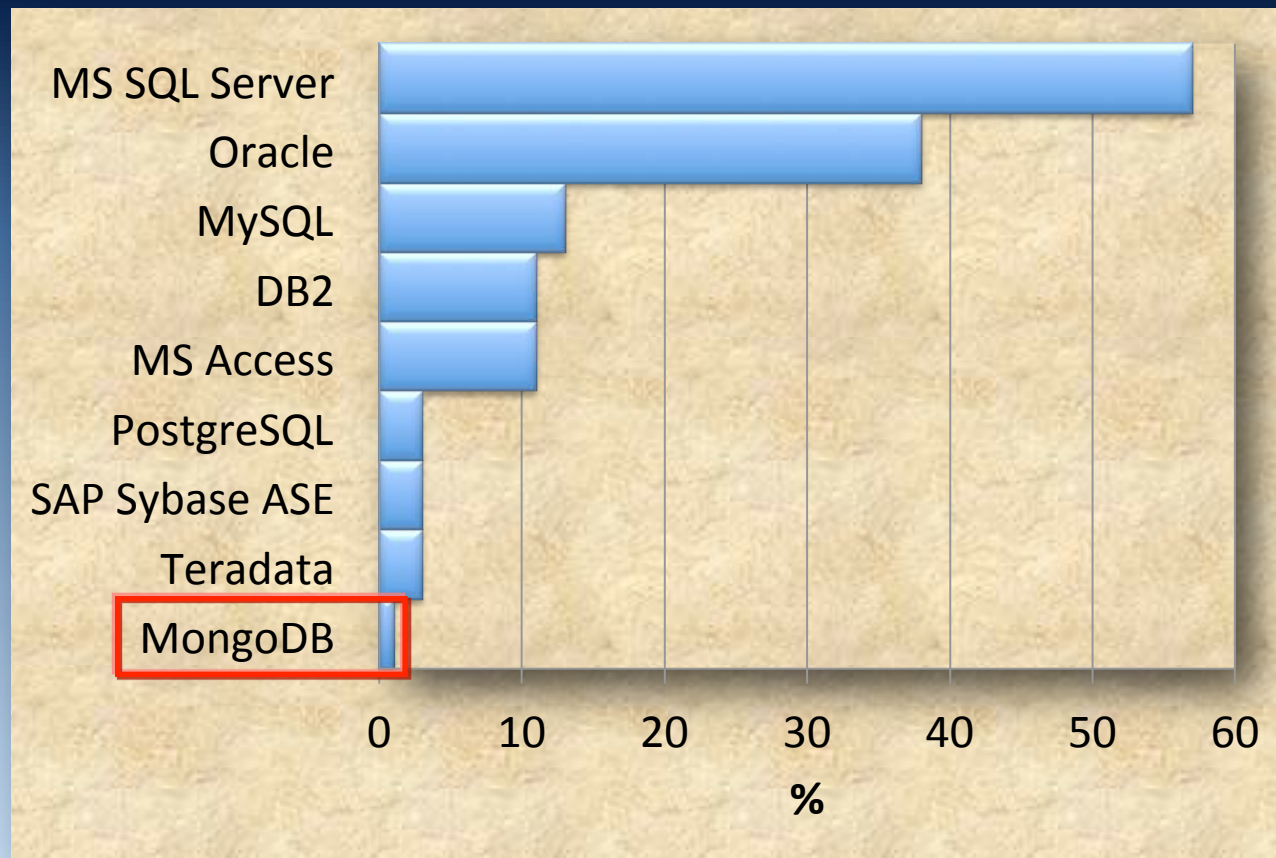


Which databases does your organization use?



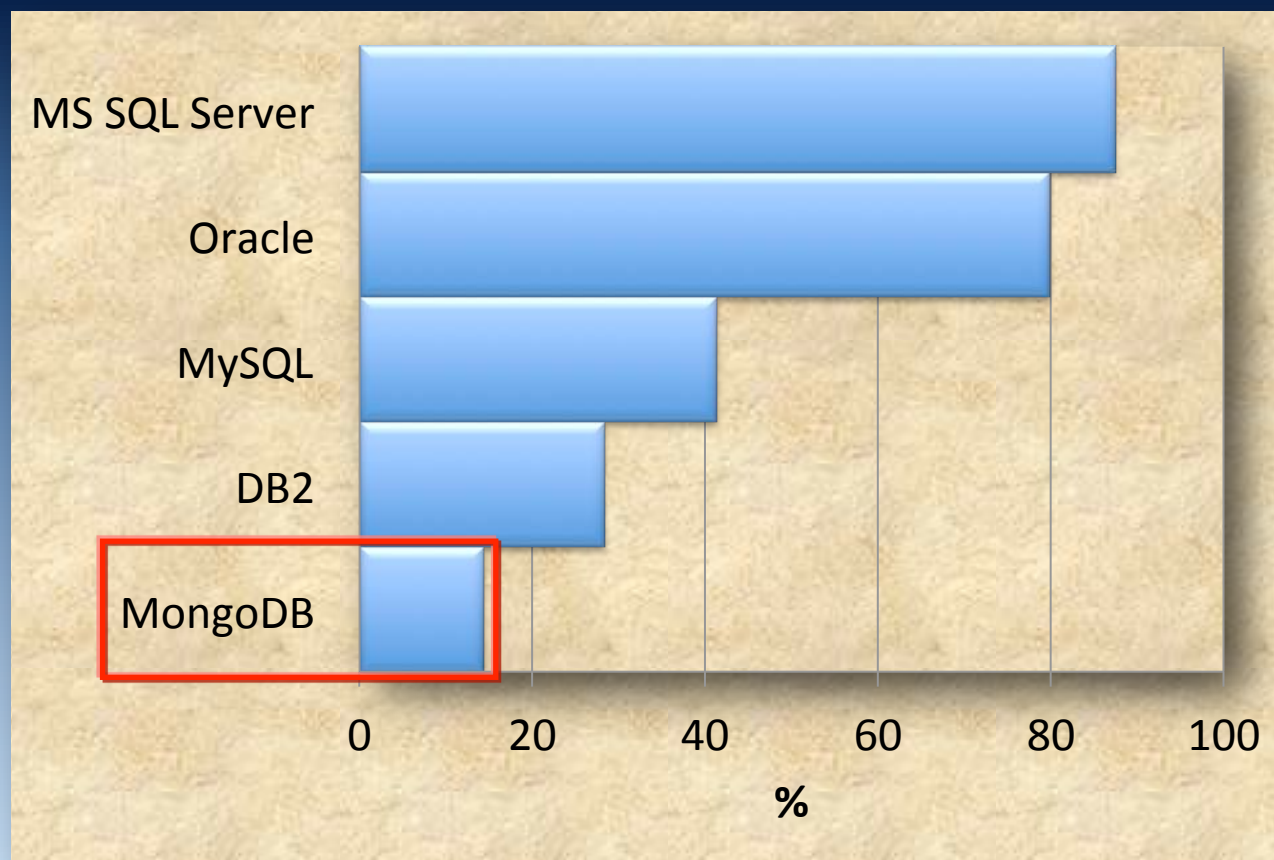
Source: "Guide to Big Data" DZone Research (2014)

Databases used for most critical functions



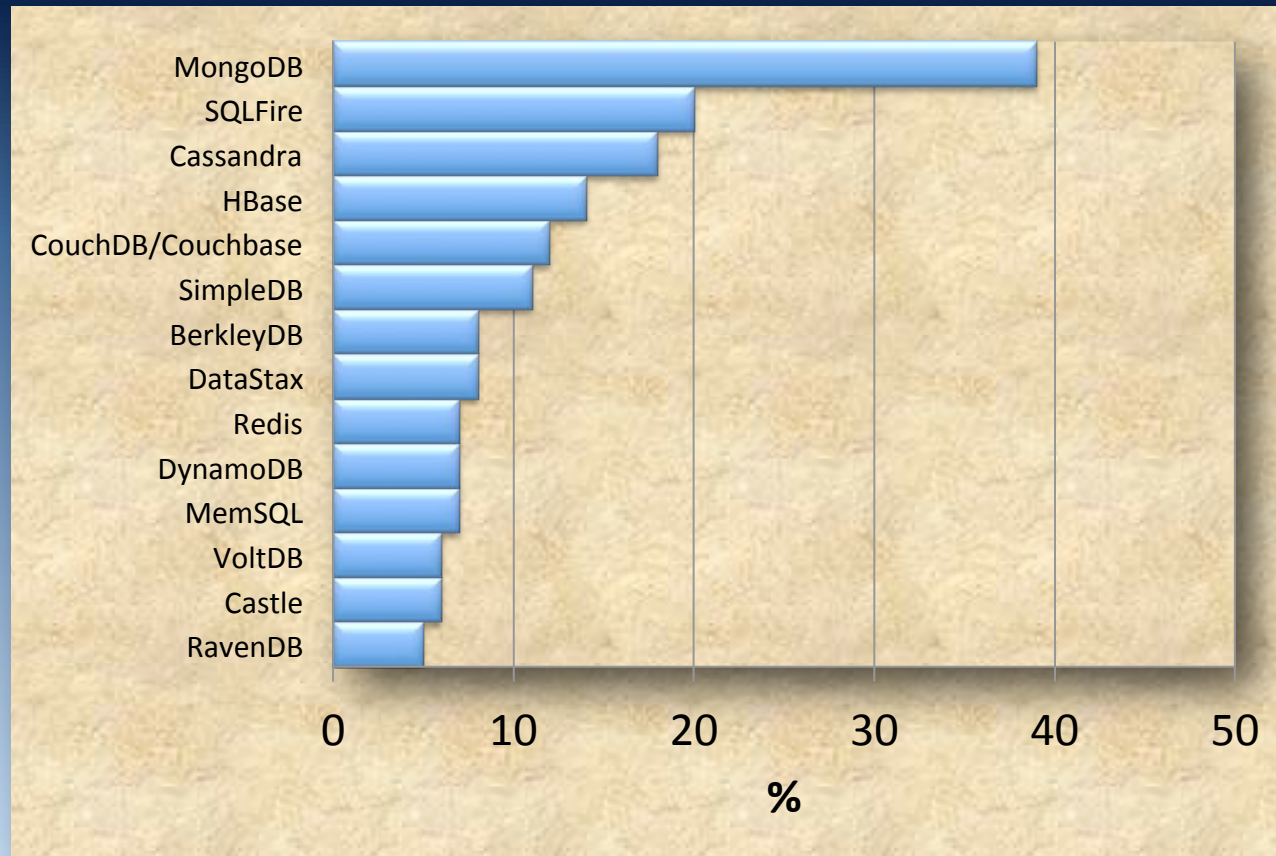
Source: "2014 State of Database Technology" InformationWeek (March 2014)

What database brands do you have running in your organization?



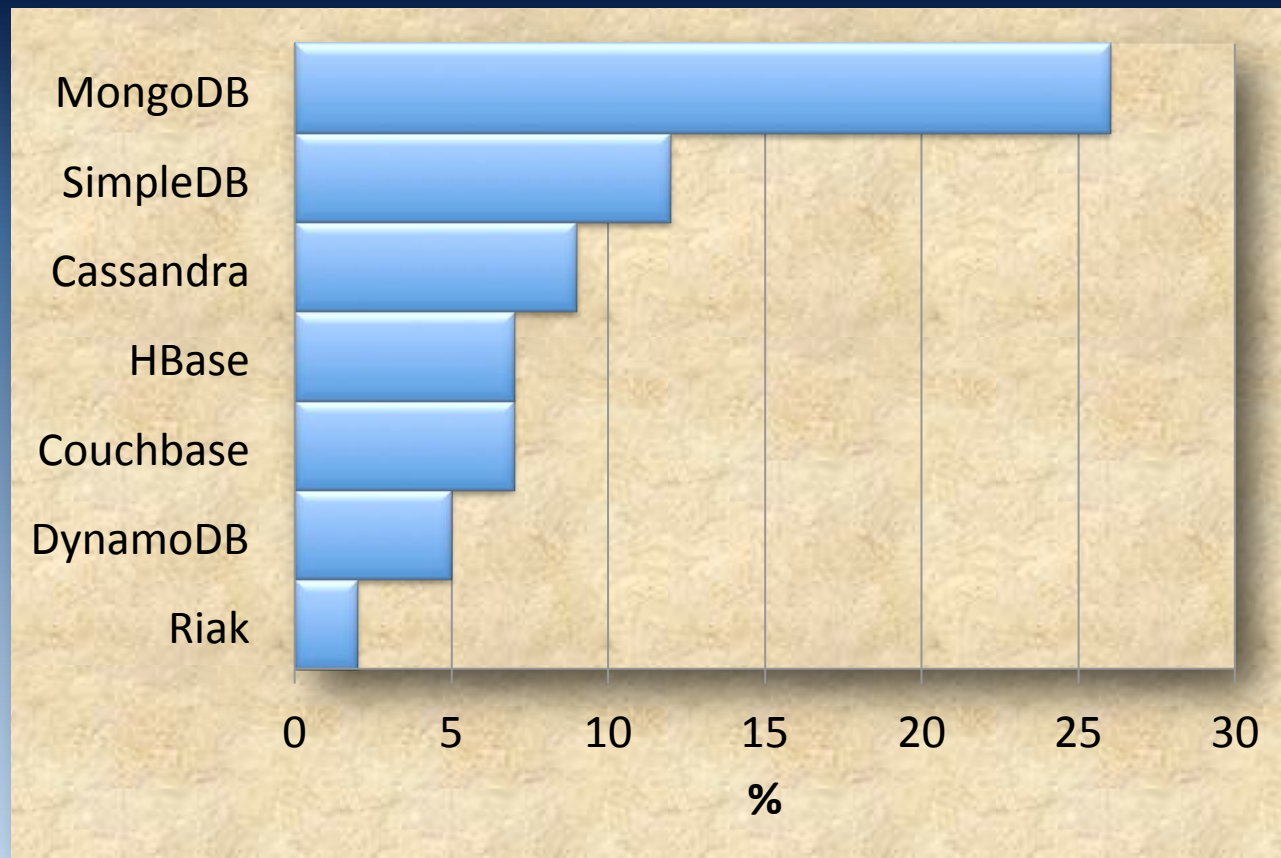
Source: "The Real World of The Database Administrator" Elliot King (March 2015)

NoSQL, NewSQL, or non-relational data store technology adoption



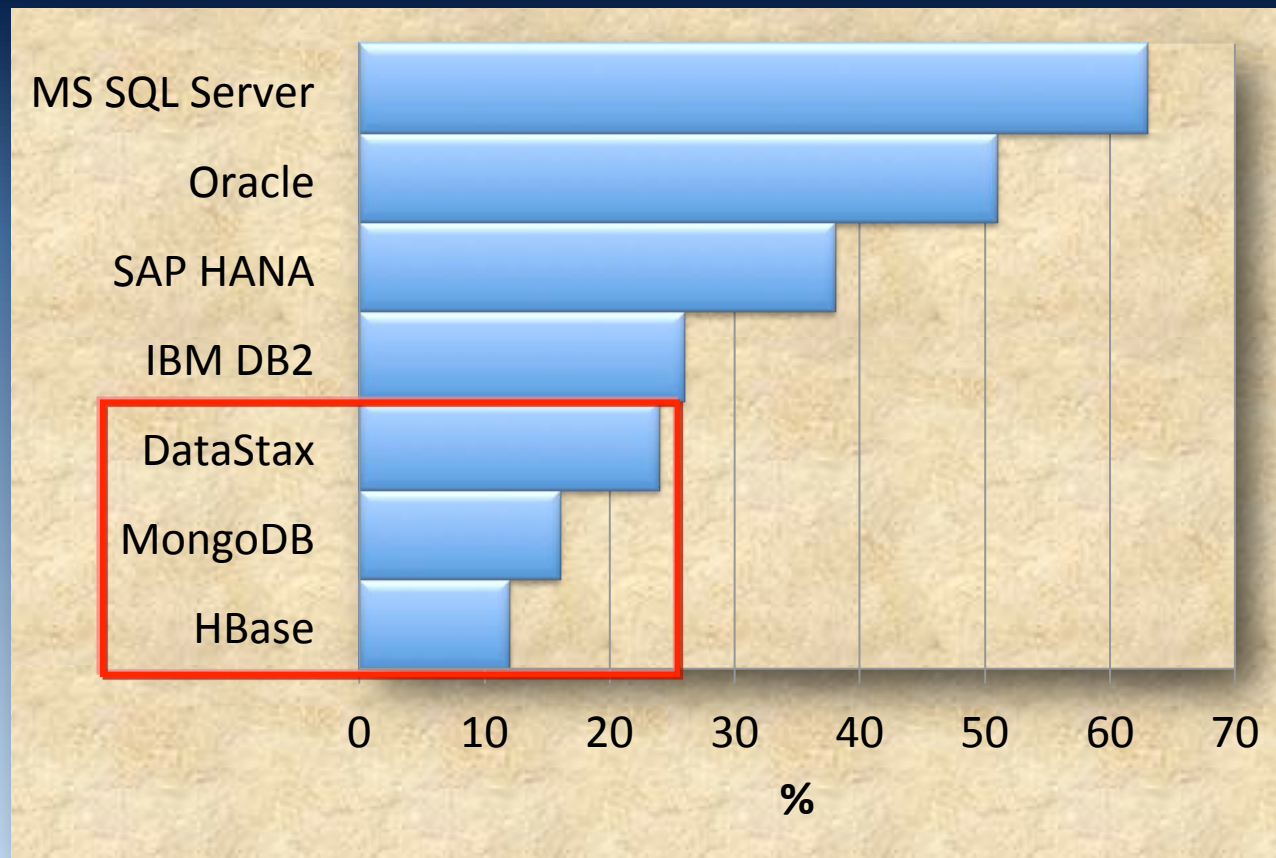
Source: "2014 Data Connectivity Outlook" Progress Software (November 2013)

NoSQL or non-relational data store technology adoption



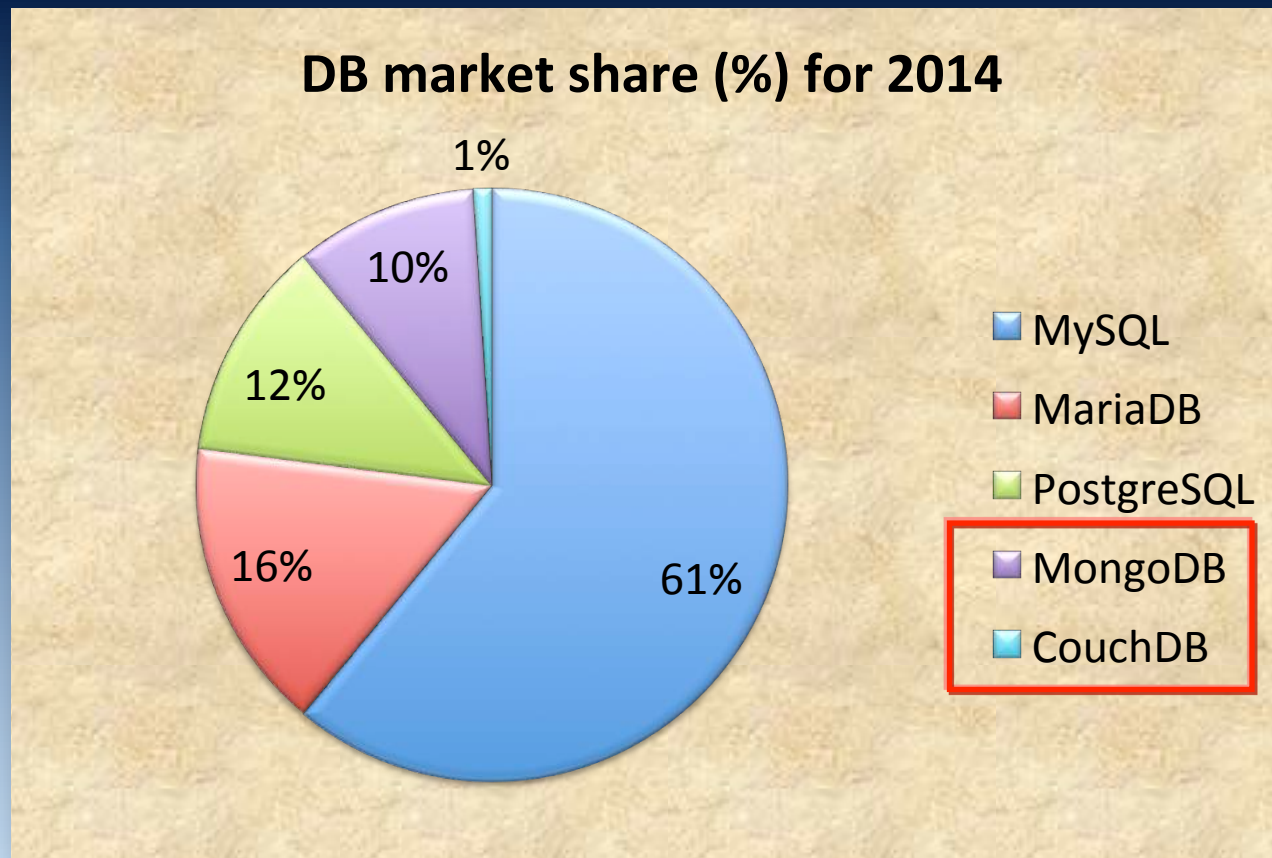
Source: "2015 Data Connectivity Outlook" Progress Software (April 2015)

When deploying new apps, which DB alternatives do you evaluate?



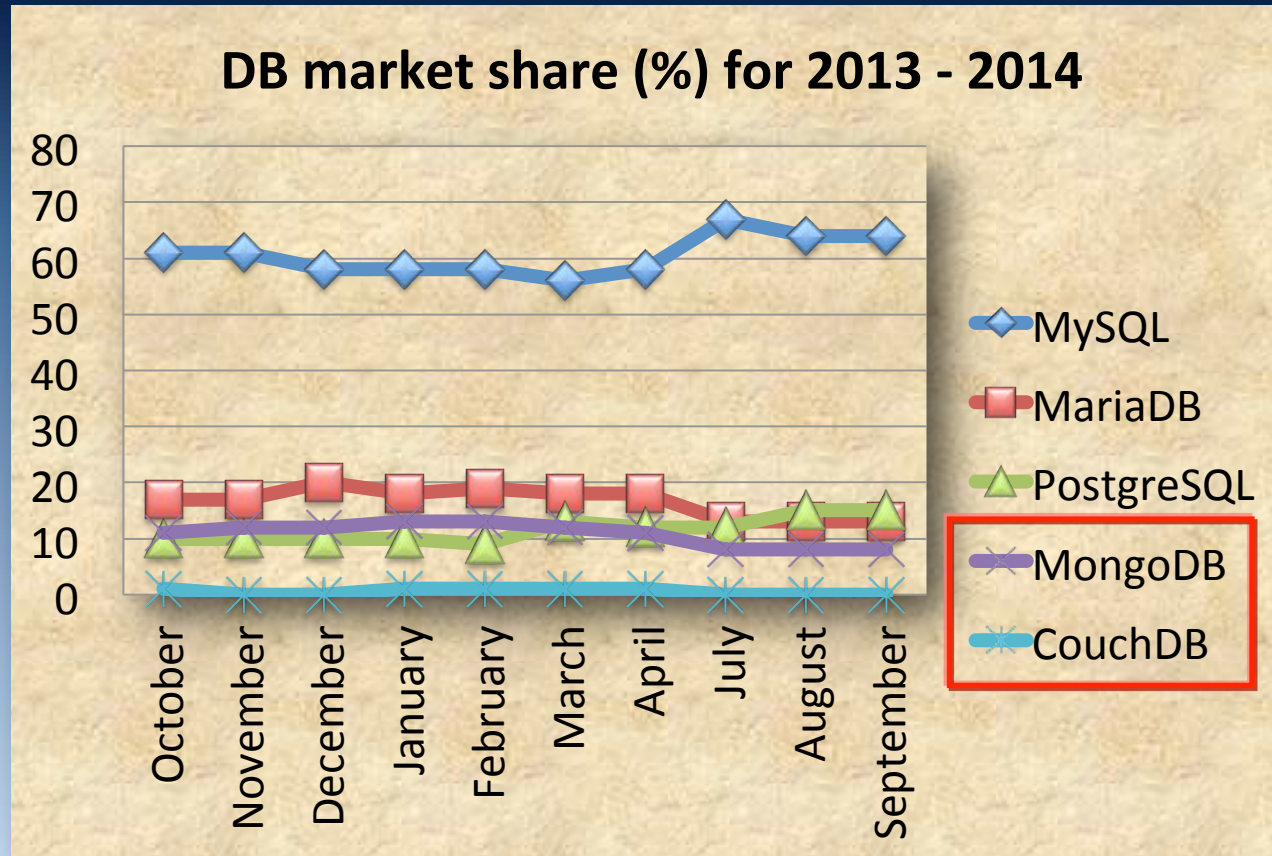
Source: Cowen and Company Mid-Year 2015 IT Spending Survey (May 2015)

Hosting example ...



Source: "Software Stacks Market Share: 2014 Summary" Tetiana Markova (13 January 2015)

Hosting example

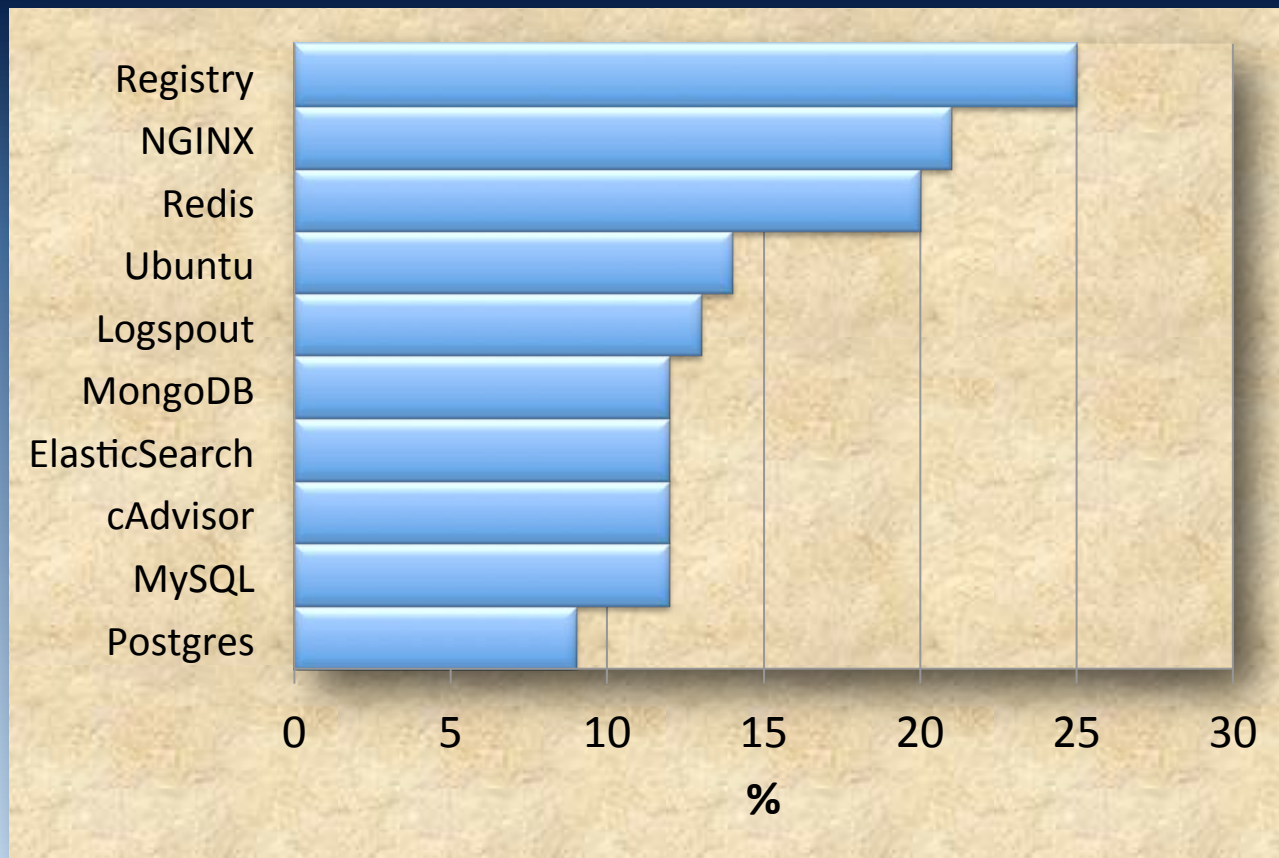


Which DB are you using or do you plan to use in your Container?



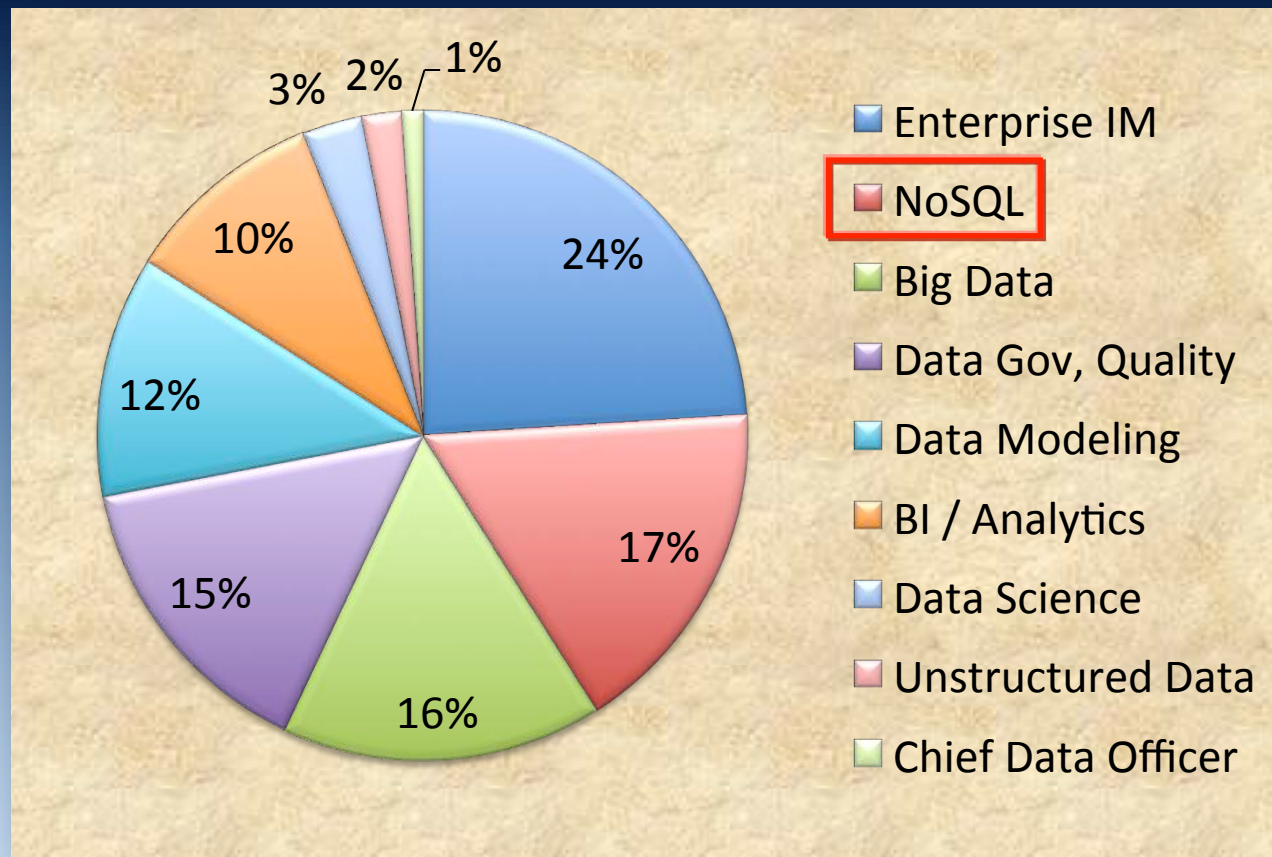
Source: "The Current State of Container Usage" ClusterHQ and DevOps.com (June 2015)

Top technologies running on Docker



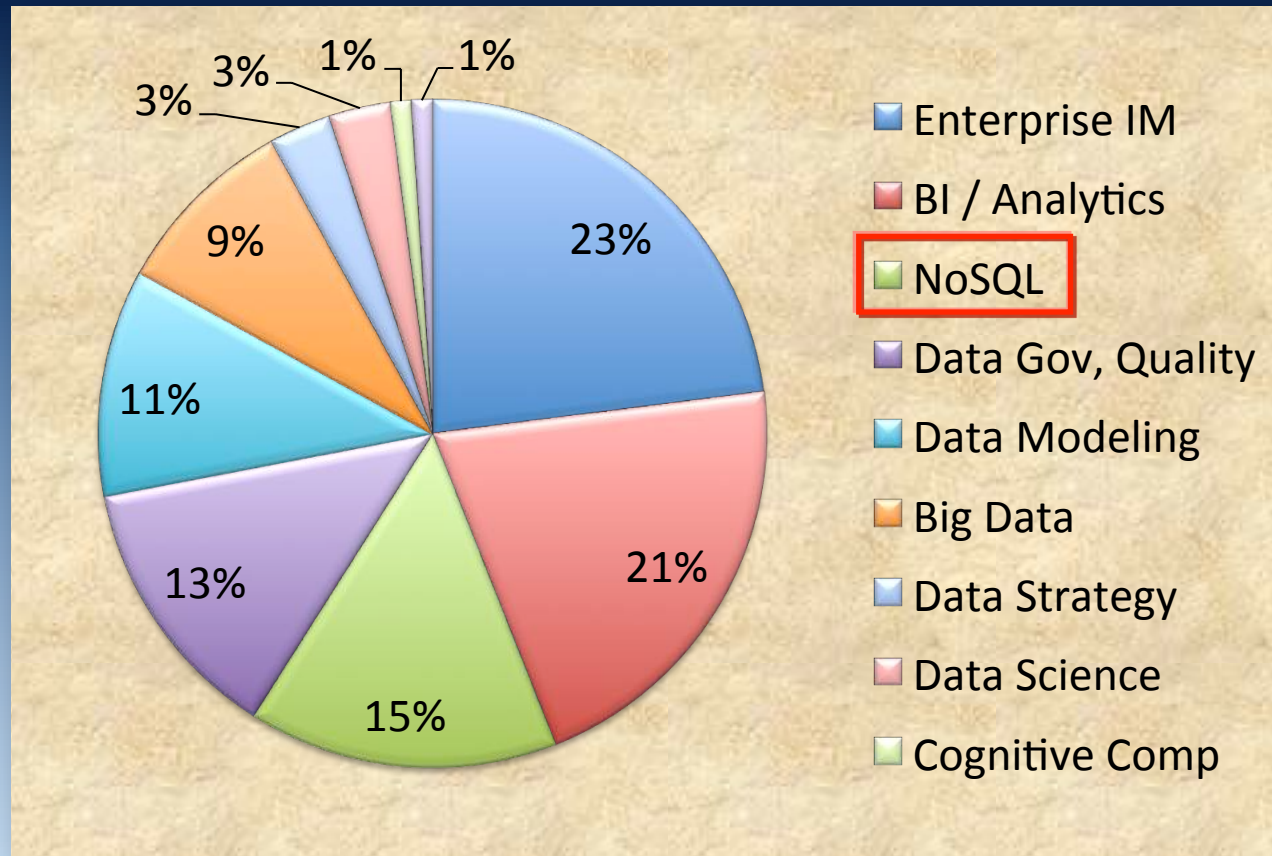
Source: "8 Surprising Facts About Real Docker Adoption" Datadog (December 2015)

Top 2013 DM topics



Source: "Top 20 Hottest Data Management Posts Year-to-Date 2014" Shannon Kempe (2 July 2014)

Top 2014 DM topics



Source: "Top 20 Hottest Data Management Posts Year-to-Date 2015" Shannon Kempe (2 July 2015)

NoSQL



STARRING AGILE DEV AND POLY GLOT

EPISODE IV: A NEW HOPE FOR DEVELOPERS

NO SQL THE MOVIE!

THE WAY DEVELOPERS REALLY THINK

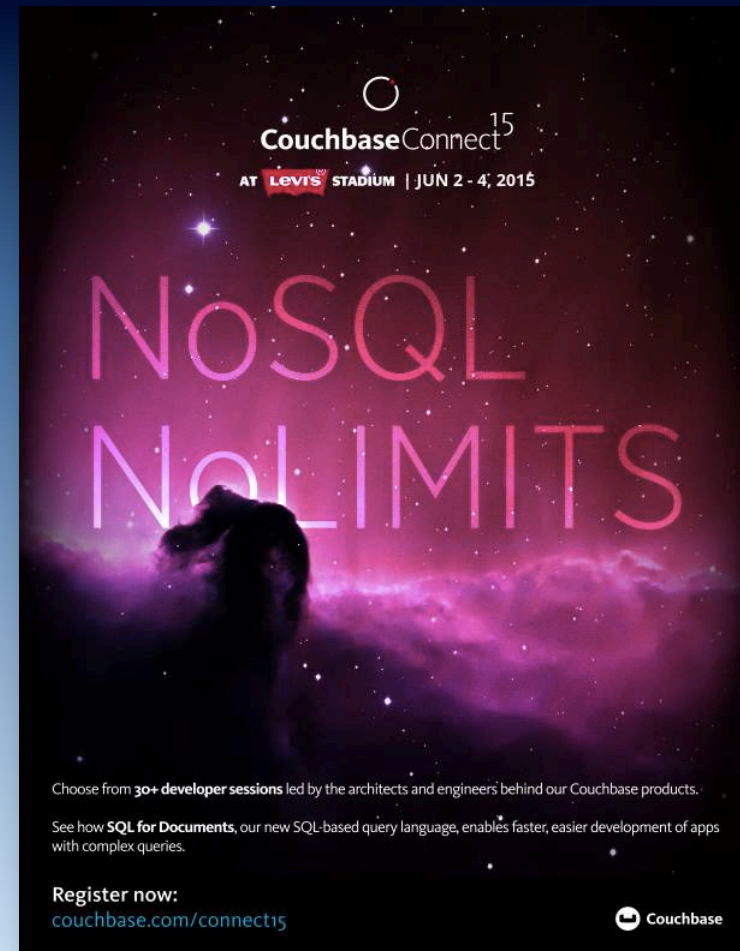
01.01.2013

DIRECTED BY AKMAL CHAUDHRI PRODUCED BY AKMAL CHAUDHRI WRITTEN BY AKMAL CHAUDHRI DISTRIBUTED BY AKMAL CHAUDHRI SOUNDTRACK BY
AKMAL CHAUDHRI MUSIC BY AKMAL CHAUDHRI EDITED BY AKMAL CHAUDHRI COPYRIGHT AKMAL CHAUDHRI

PG PARENTAL GUIDANCE SUGGESTED
SOME MATERIAL MAY NOT BE SUITABLE FOR CHILDREN



Imitation is the sincerest form of flattery - thank you Couchbase!



“The Stars, Like Dust”



... a squadron of small, flitting ships that had struck and vanished, then struck again, and made scrap of the lumbering titanic ships that had opposed them ... abandoning power alone, stressed speed and co-operation ...

-- Isaac Asimov

NoSQL The Movie!

~~Sequel~~



History in No-tation

1970: NoSQL = We have no SQL

1980: NoSQL = Know SQL

2000: NoSQL = No SQL!

2005: NoSQL = Not only SQL

2013: NoSQL = No, SQL!

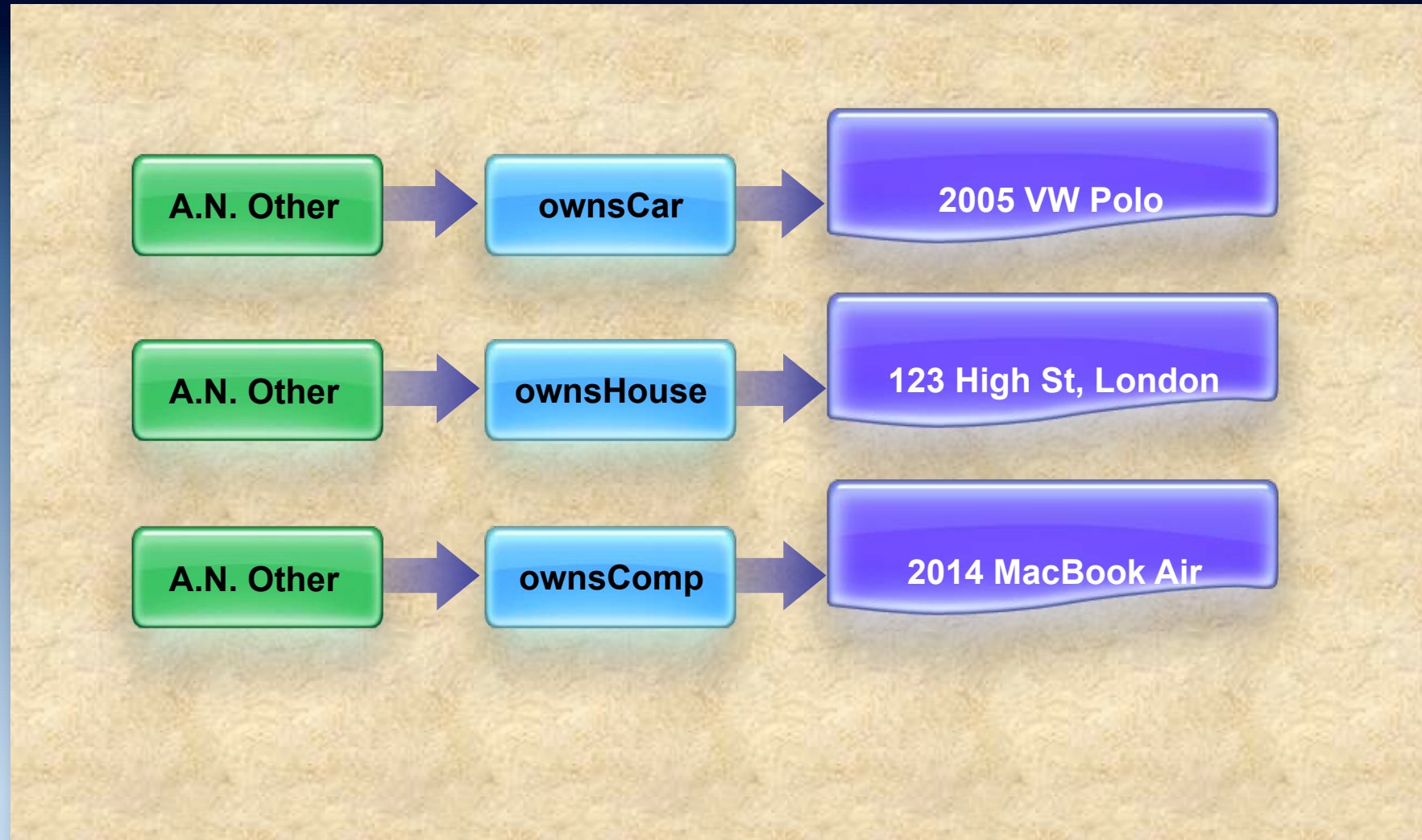
The meme changed



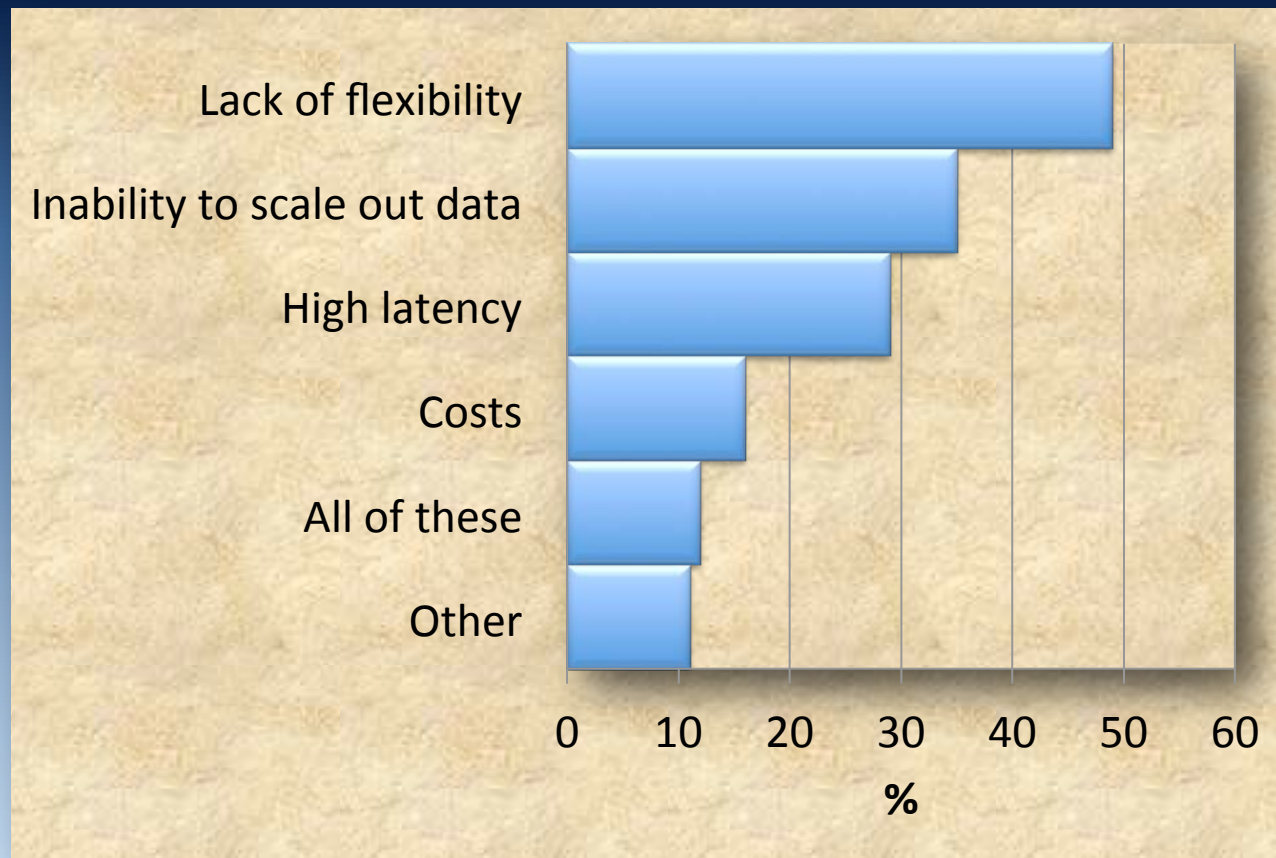
Why did NoSQL datastores arise?

- Some applications need very few database features, but need high scale
- Desire to avoid data/schema pre-design altogether for simple applications
- Need for a low-latency, low-overhead API to access data
- Simplicity - do not need fancy indexing - just fast lookup by primary key

Scenario where NoSQL is useful

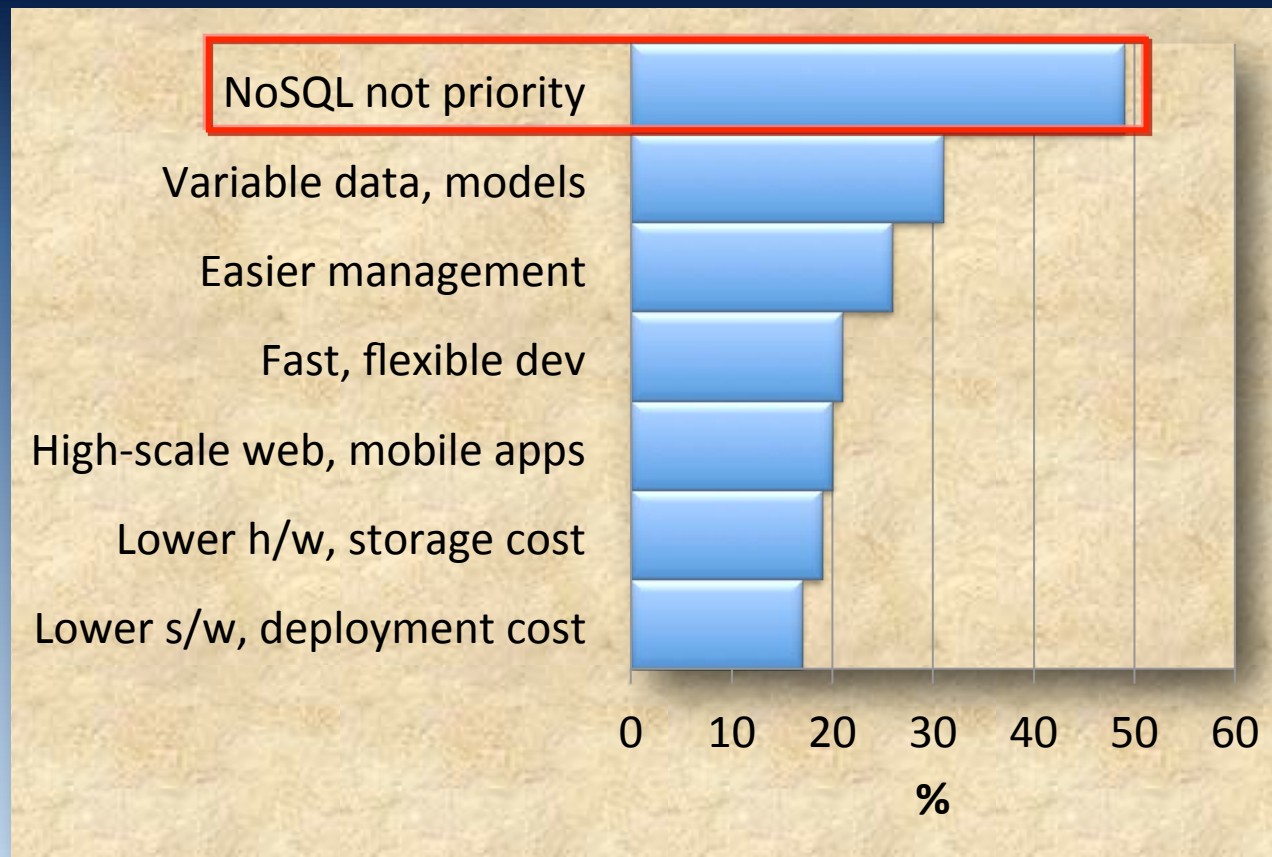


What is the biggest DM problem driving your use of NoSQL?



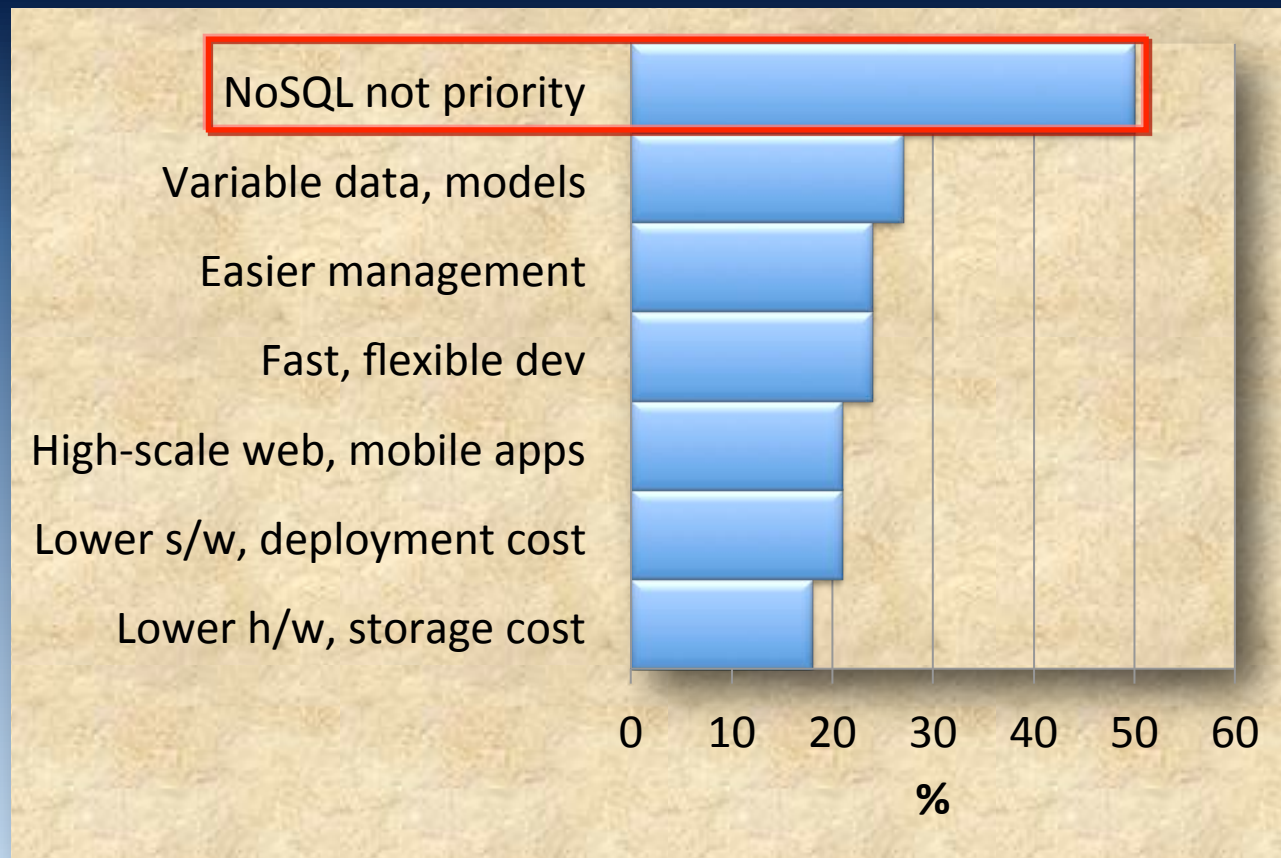
Source: Couchbase NoSQL Survey (December 2011)

Eye on NoSQL 2013



Source: "2014 Analytics, BI, and Information Management Survey" InformationWeek (November 2013)

Eye on NoSQL 2014



Source: "2015 Analytics & BI Survey" InformationWeek (December 2014)

Schema-free



Source: Shutterstock Image ID 128628794

But ...



We started using mongo early 2009, and even just one year out it feels so much more painful to maintain than our Postgres or MySQL systems that have been around since 1999! My theory is that NoSQL sacrifices maintenance and future development effort for the sake of startup development.

-- Luke Crouch

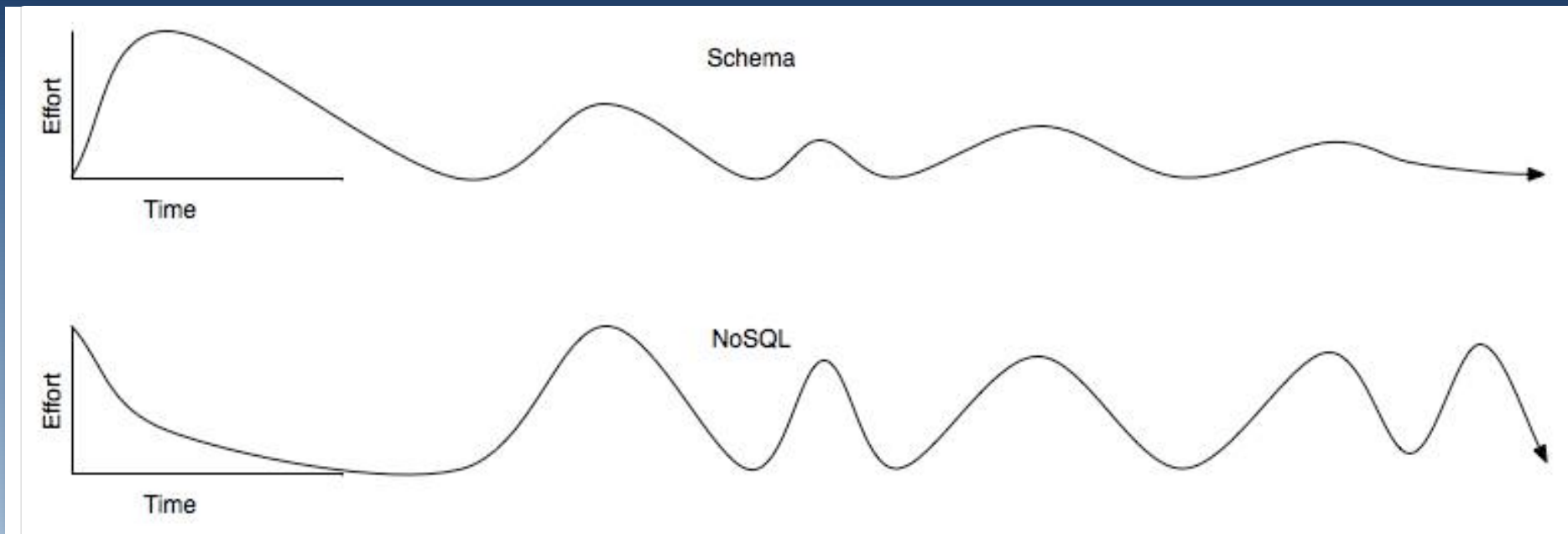
And ...



Inquiries from Gartner clients indicate that schema design for NoSQL DBMSs is one of the biggest barriers to adopting this new technology. Simply selecting a NoSQL DBMS and hoping the underlying technology will accommodate poor design choices will lead to a poorly performing application and database, and to rework.

-- Adam M. Ronthal and Nick Heudecker

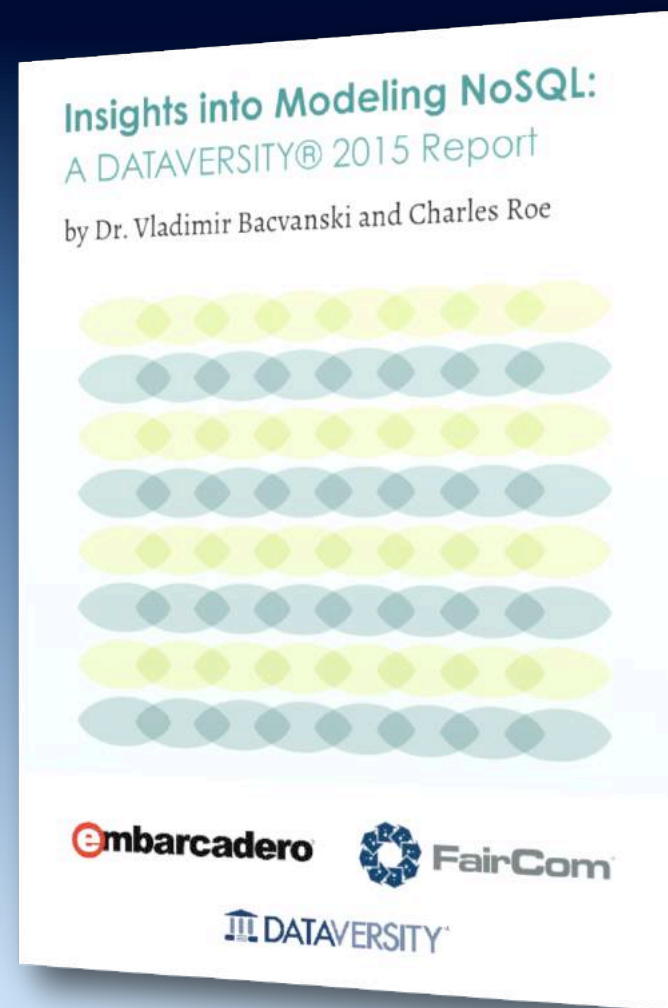
Schema



Source: Luke Crouch, used with permission

Data modelling

- 32% do not do data modelling for their NoSQL system, they simply code the application
- 46% of the data modelling with NoSQL is done by the programmer who uses the NoSQL store



Big data

Variety



Velocity



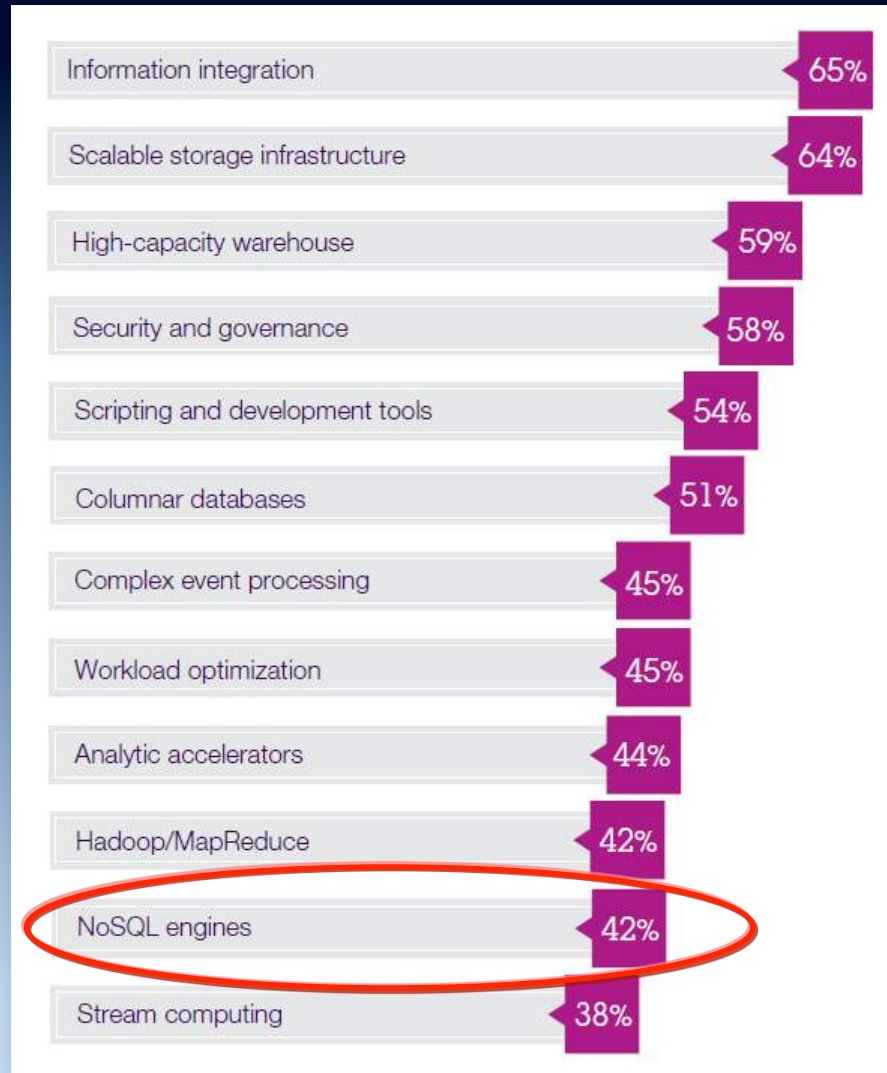
Volume



What is Big Data?

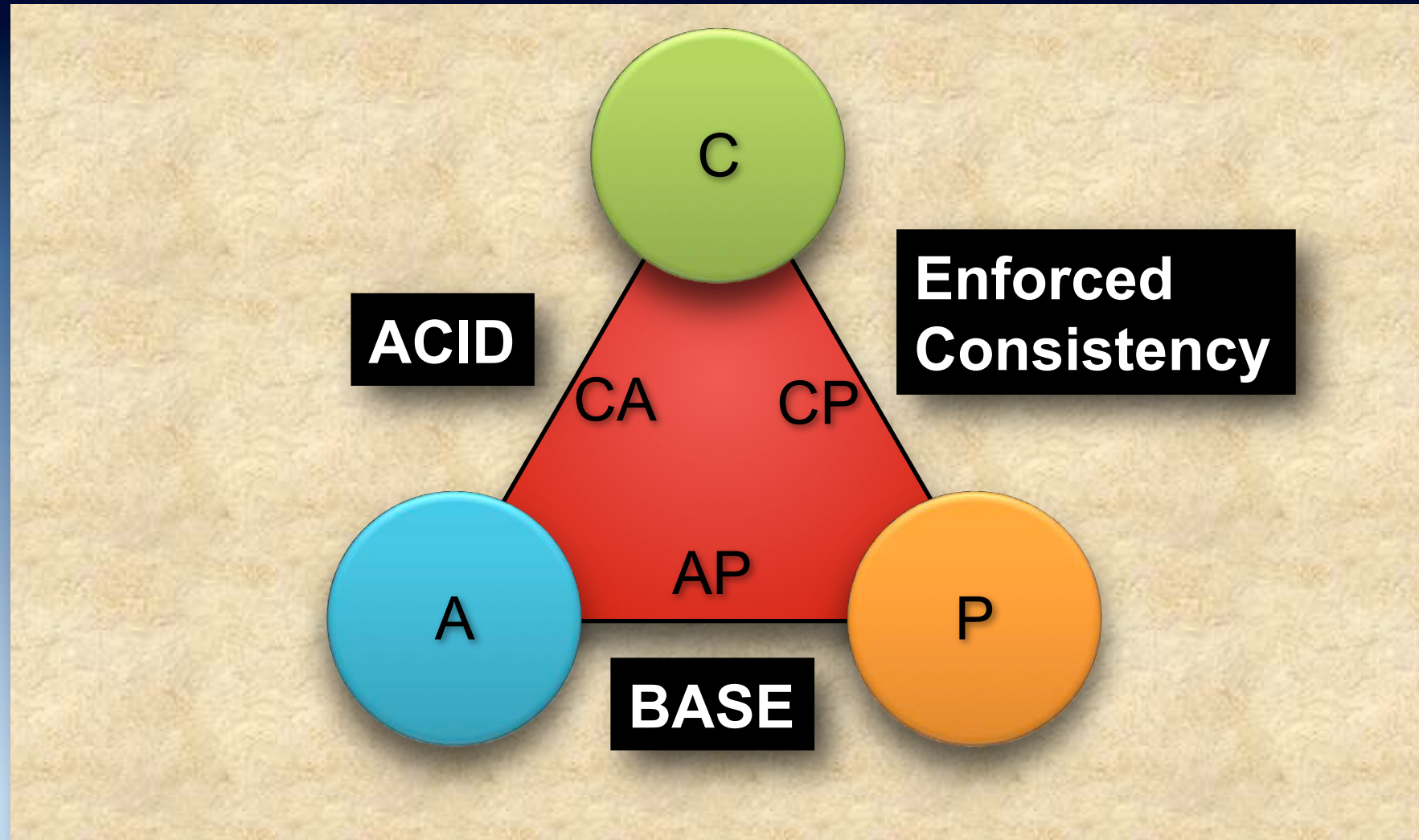
Byte	: One grain of rice	Hobbyist
Kilobyte	: Cup of rice	
Megabyte	: 8 bags of rice	Desktop
Gigabyte	: 3 semi trucks	
Terabyte	: 2 container ships	Internet
Petabyte	: Blankets Manhattan	
Exabyte	: Blankets west coast states	Big Data
Zettabyte	: Fills the Pacific Ocean	
Yottabyte	: Earth size rice ball	

Big data infrastructure



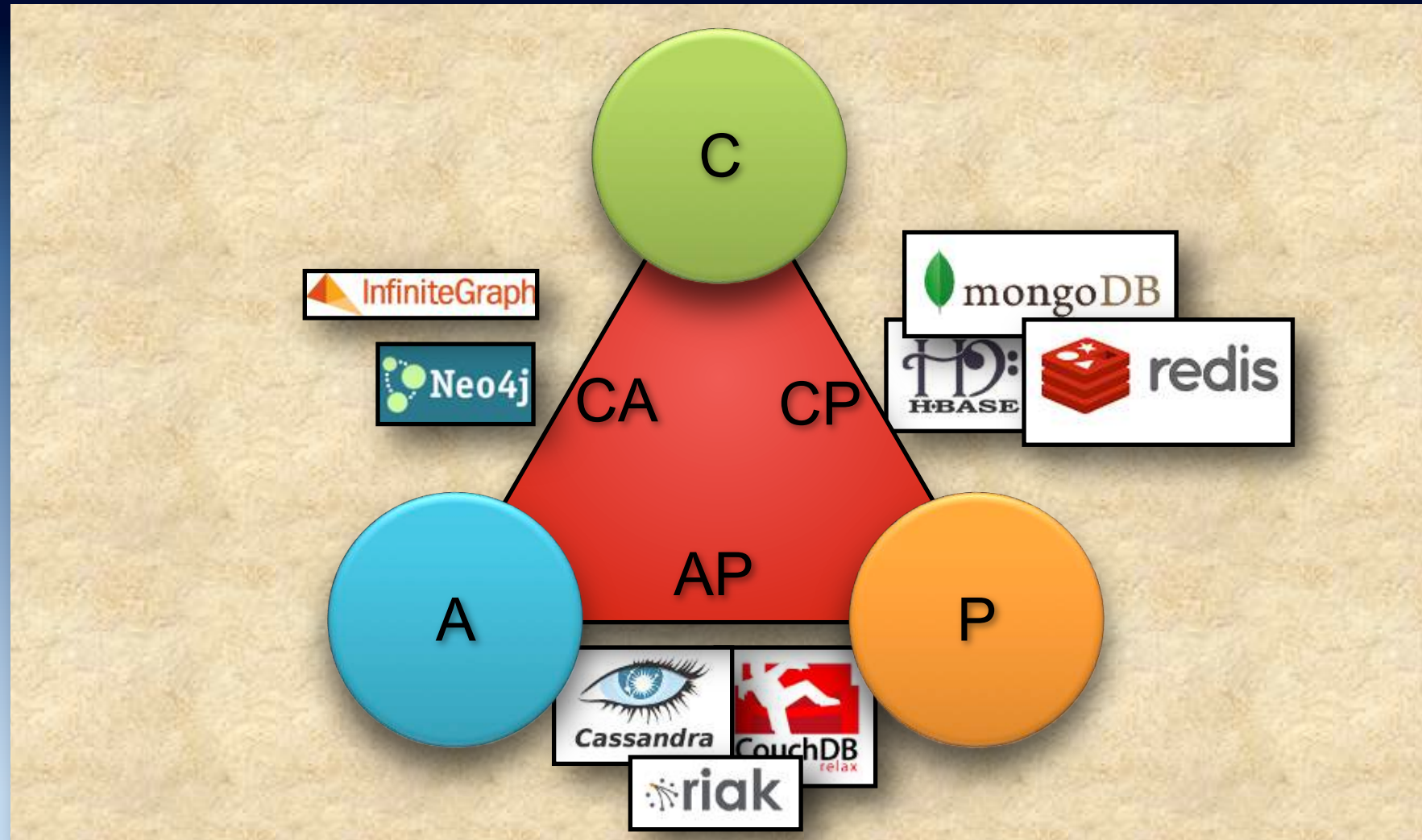
Source: "Analytics: The real-world use of big data" IBM and University of Oxford (October 2012)

Brewer's CAP "Theorem" ...



Source: After <http://guide.couchdb.org/editions/1/en/consistency.html>

Brewer's CAP "Theorem"



ACID vs. BASE ...



- Atomicity
- Consistency
- Isolation
- Durability



- Basically Available
- Soft state
- Eventual consistency

ACID vs. BASE

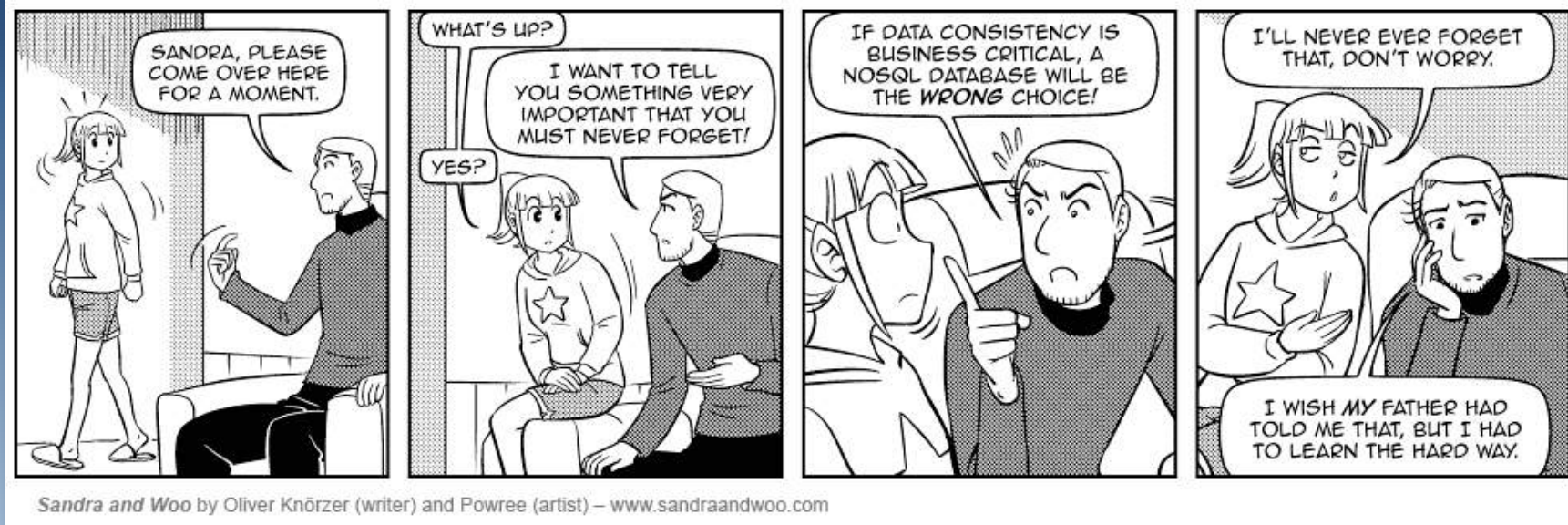
ACID	BASE
<ul style="list-style-type: none">• Strong consistency• Isolation• Focus on “commit”• Nested transactions• Conservative (pessimistic)• Availability• Difficult evolution	<ul style="list-style-type: none">• Weak consistency• Availability first• Best effort• Approximate answers OK• Aggressive (optimistic)• Simpler, faster• Easier evolution

But ...



... we find developers spend a significant fraction of their time building extremely complex and error-prone mechanisms to cope with eventual consistency and handle data that may be out of date. We think this is an unacceptable burden to place on developers and that consistency problems should be solved at the database level.

Use the right tool



Tuneable CAP

- Examples
 - Cassandra
 - MongoDB
 - Riak

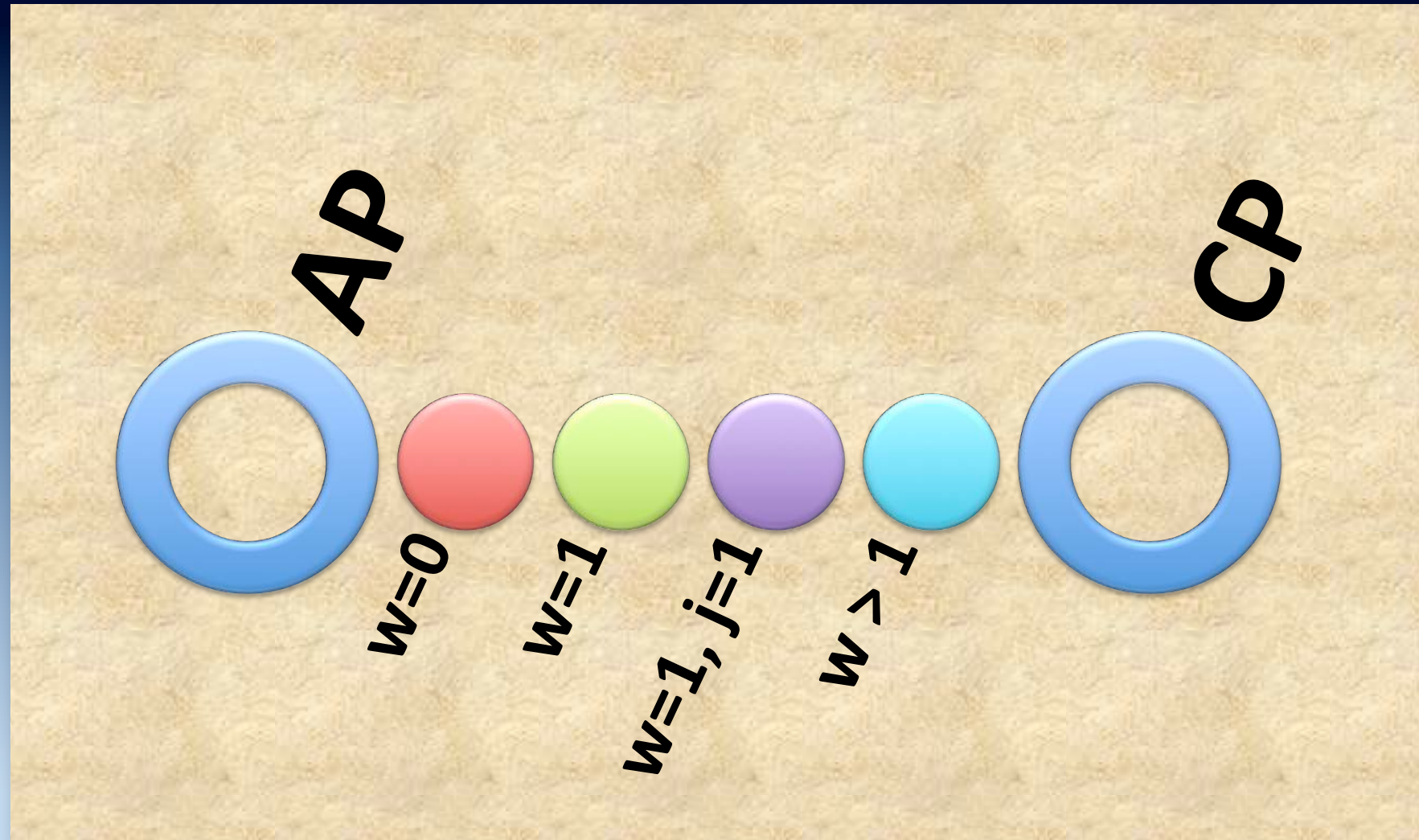


MongoDB speed vs. safety

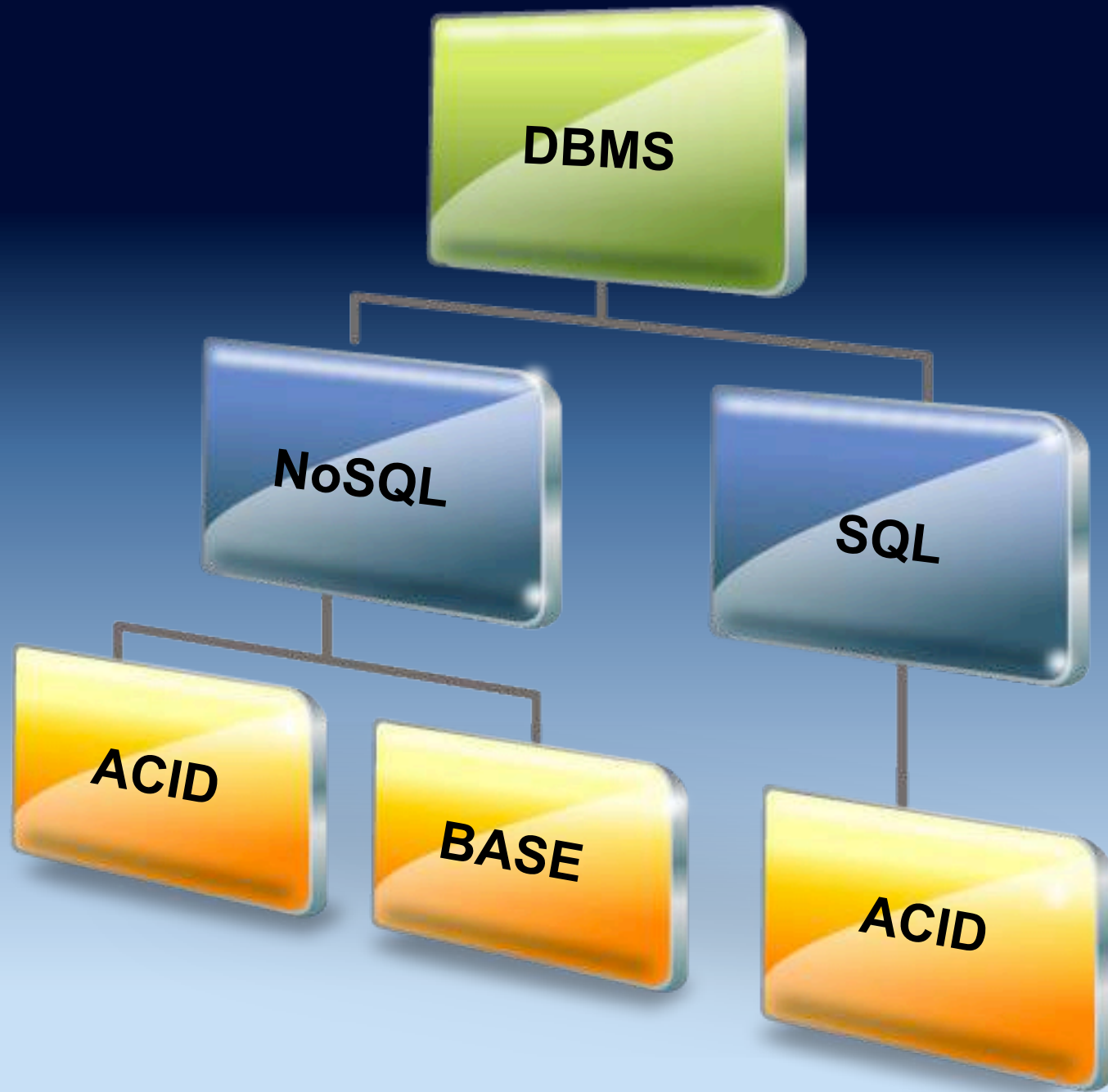
Options	WriteConcern	Notes
w=0, j=0	UNACKNOWLEDGED	Fire and Forget
w=1, j=0	ACKNOWLEDGED	Operation completed successfully in memory
w=1, j=1	JOURNALED	Operation written to the journal file
w=1, fsync=true	FSYNCED	Operation written to disk
w=2, j=0	REPLICA_ACKNOWLEDGED	Ack by primary and at least one secondary
w=majority, j=0	MAJORITY	Ack by the majority of nodes

Source: "MongoDB Replication" Philipp Krenn (30 November 2014)

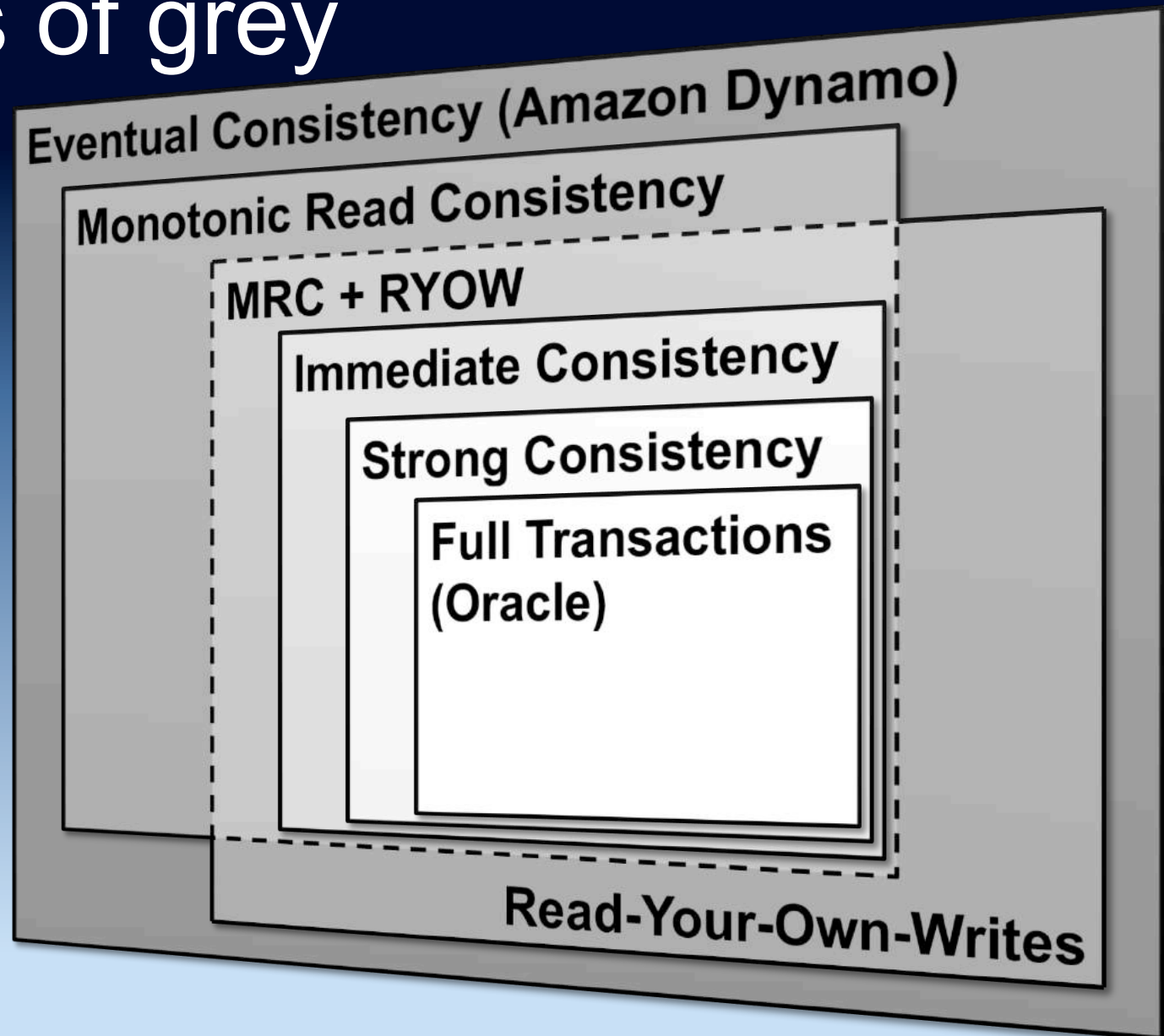
MongoDB Replica Sets



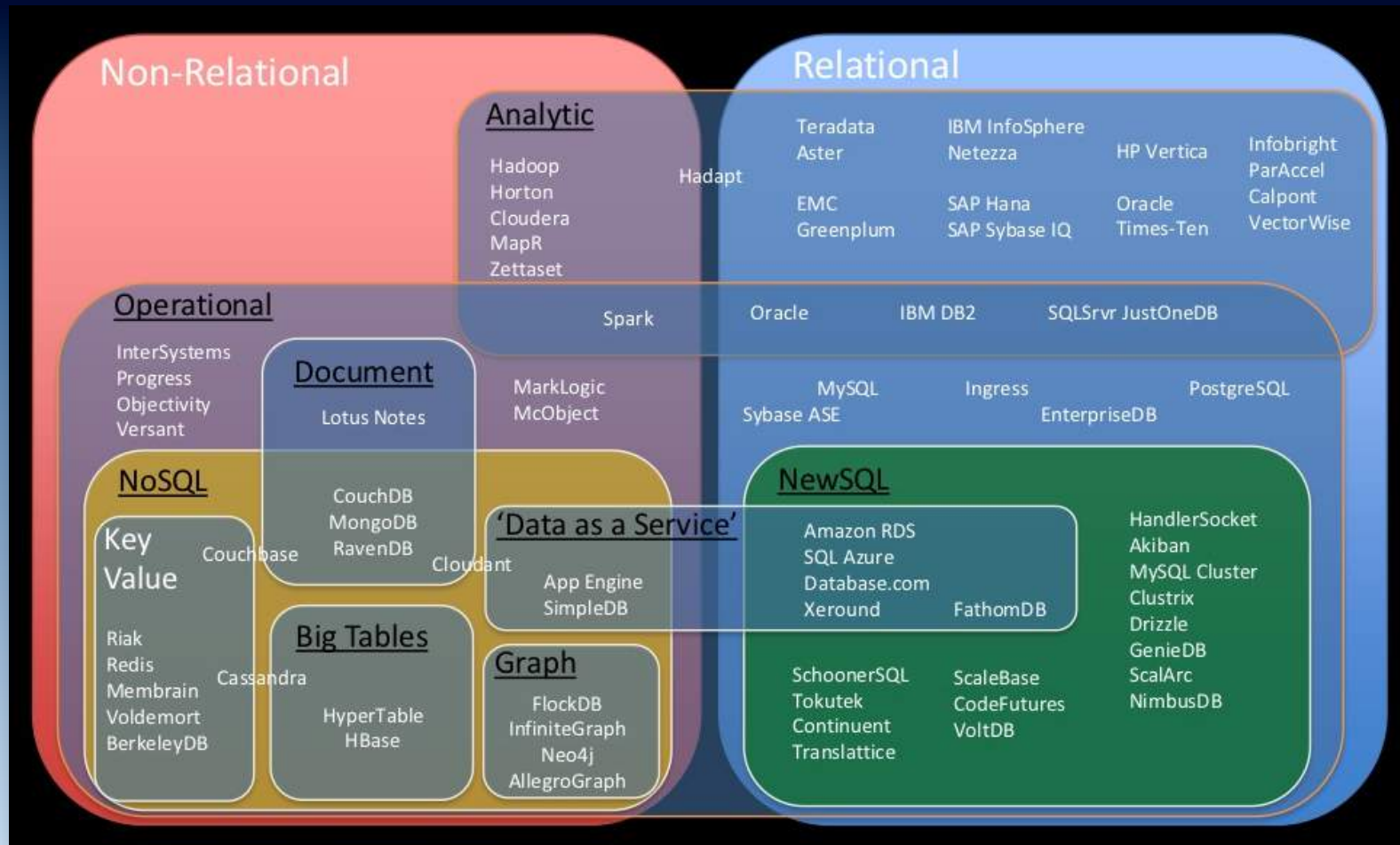
Source: Adapted from "Don't fight MongoDB" Mirko Bonadei (13 December 2013)



Shades of grey



Choices, choices



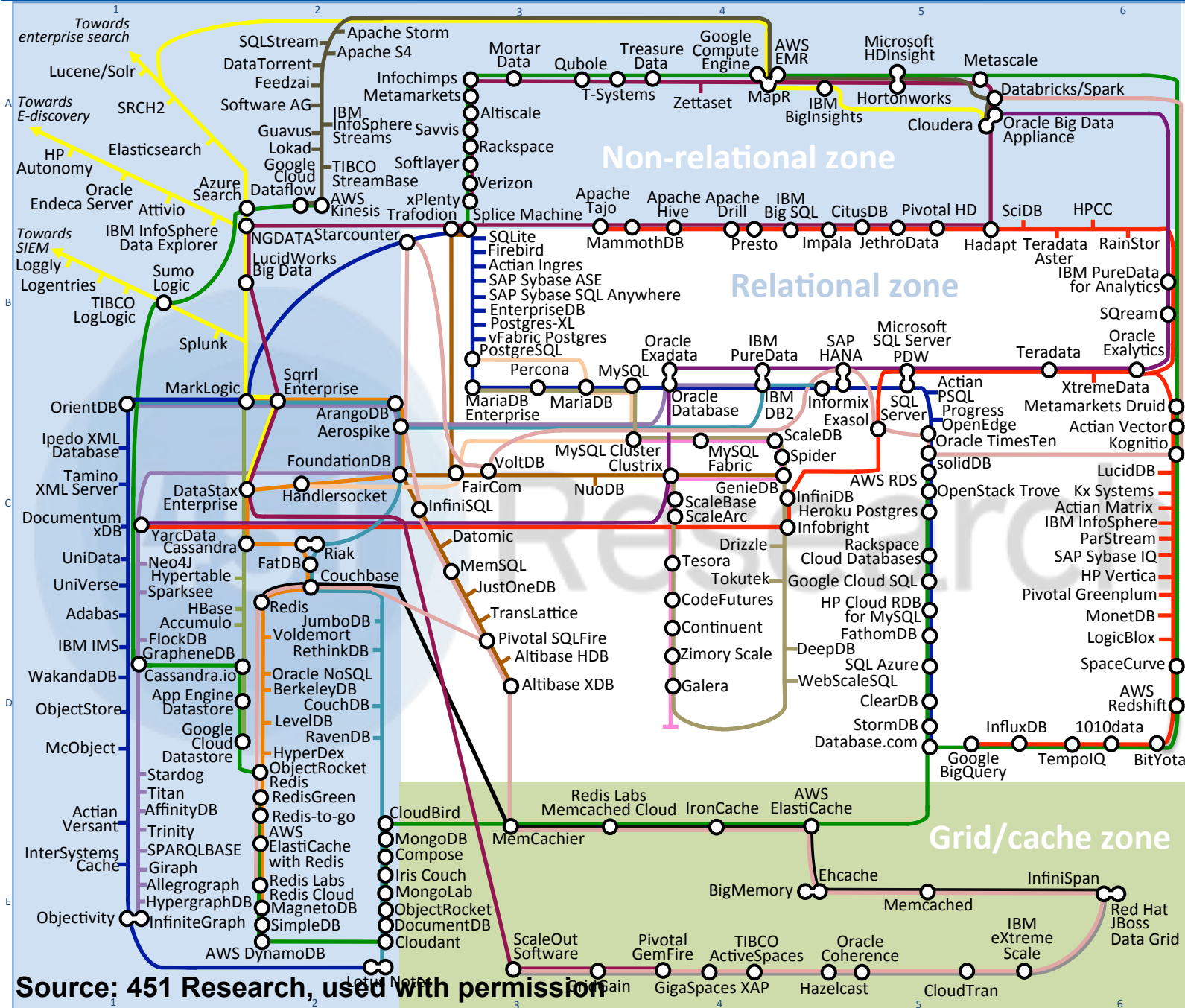
Source: Infochimps, used with permission

451 Research: Data Platforms Landscape Map – September 2014



© 2014 by 451 Research LLC. All rights reserved

- Key:**
- General purpose
 - Specialist analytic
 - -as-a-Service
 - BigTables
 - Graph
 - Document
 - Key value stores
 - Key value direct access
 - Hadoop
 - MySQL ecosystem
 - Advanced clustering/sharding
 - New SQL databases
 - Data caching
 - Data grid
 - Search
 - Appliances
 - In-memory
 - Stream processing



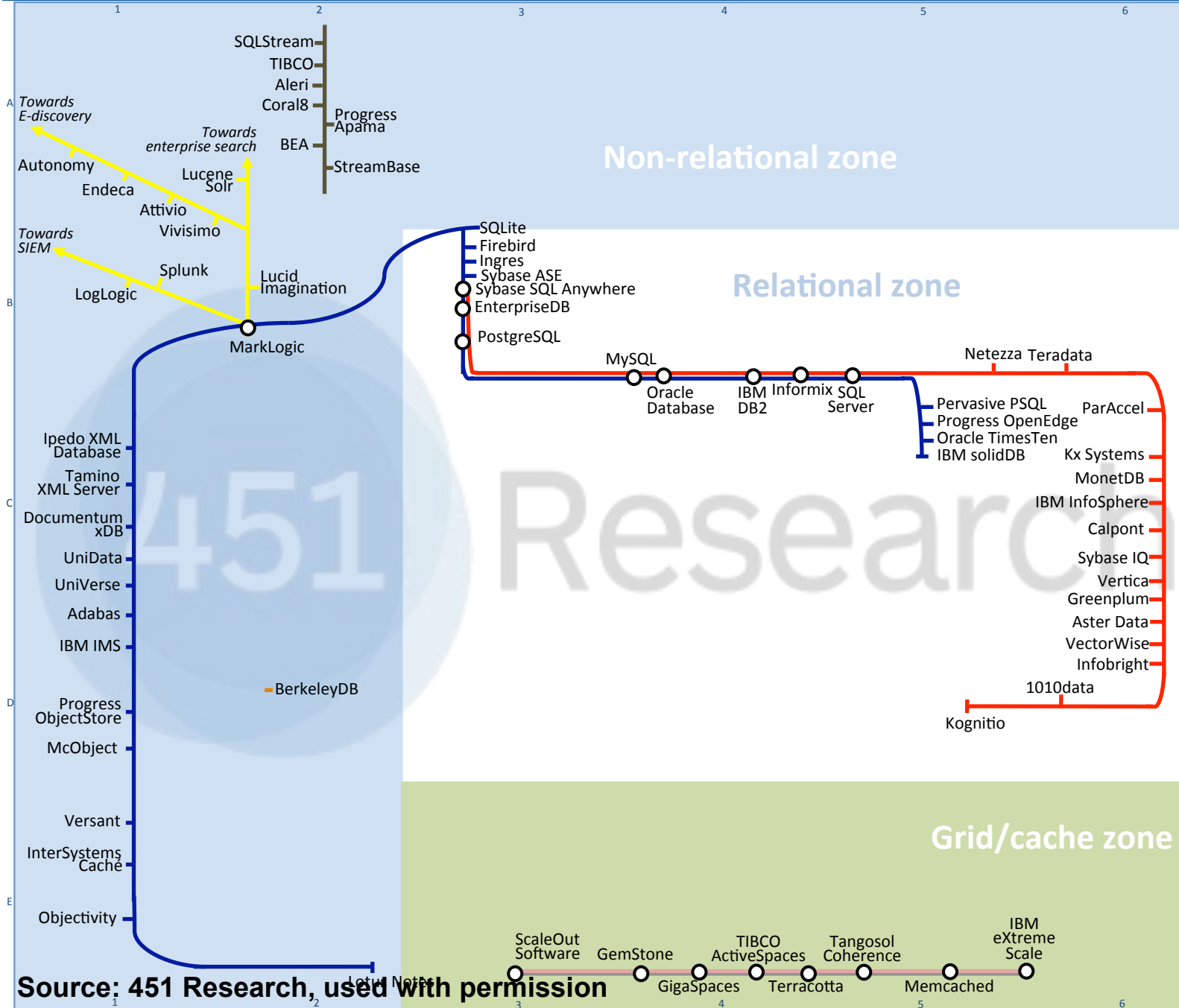
Source: 451 Research, used with permission

451 Research: Data Platforms Landscape Map – ~2009



© 2014 by 451 Research LLC. All rights reserved

- Key:**
- General purpose
 - Specialist analytic
 - Data grid/cache
 - Search
 - In-memory
 - Stream processing



Source: 451 Research, used with permission

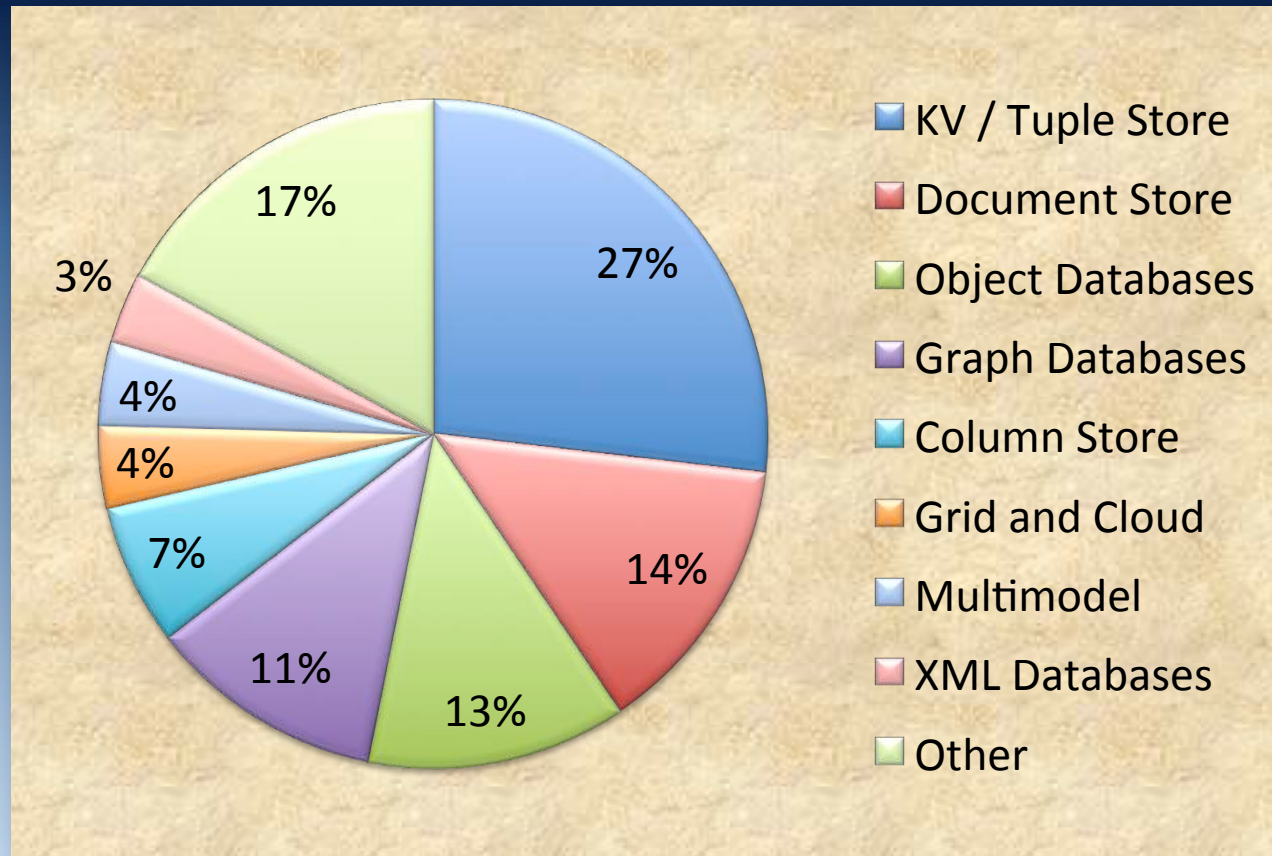
How many systems? ...



There are a lot of Key/Value stores and distributed schema-free Document Oriented Databases out there. They're springing up like weeds in a spring garden. And folks love to blog about them and/or talk about how their favorite is better than the others (or MySQL).









-- Jeremy Zawodny

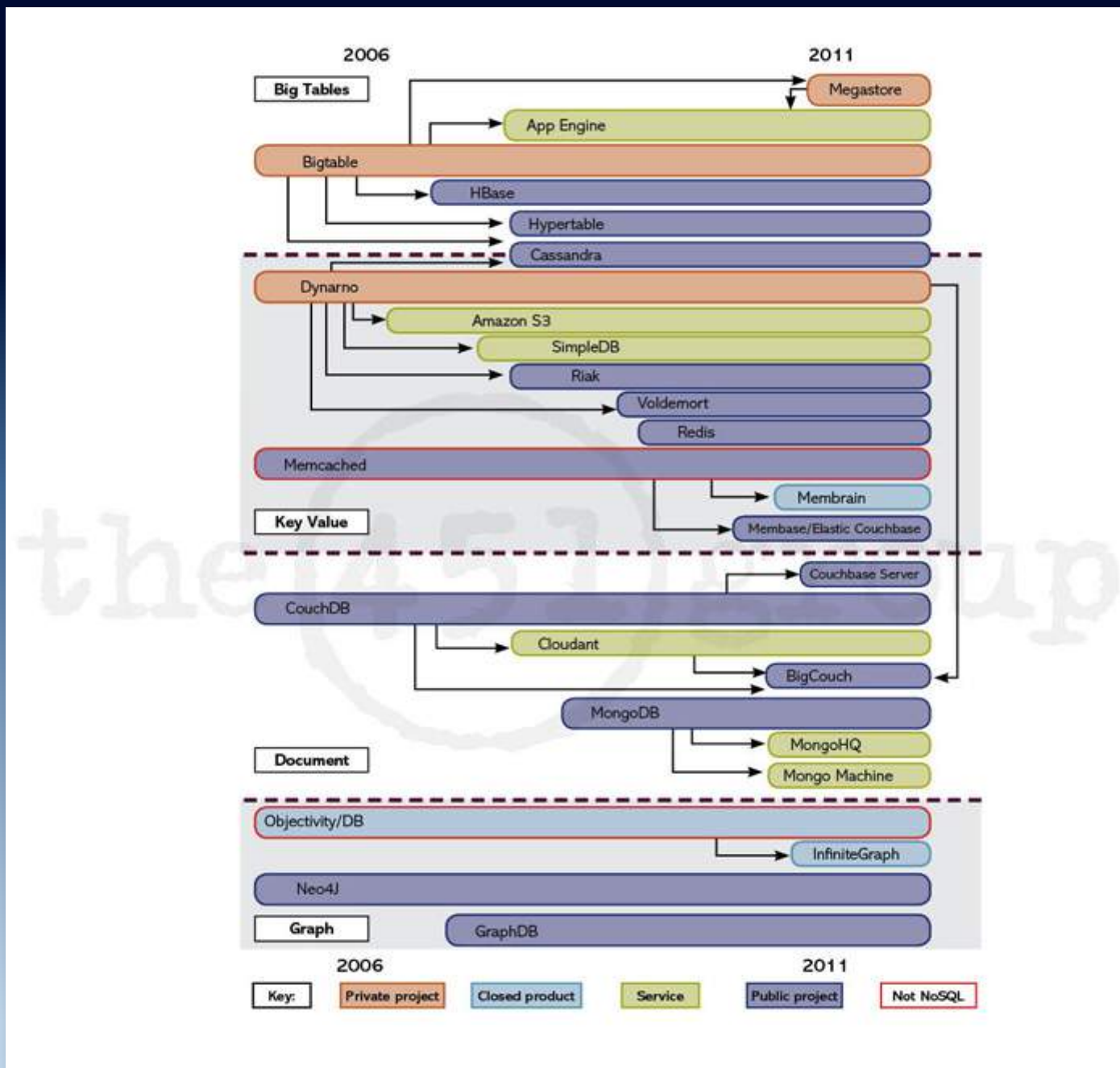
How many systems?



Source: <http://nosql-database.org/> (24 March 2015)

Major categories of NoSQL ...

Type	Examples
Key-Value store	 redis  riak
Column store	 Cassandra  HBASE
Document store	 CouchDB  mongoDB
Graph store	 InfiniteGraph  Neo4j



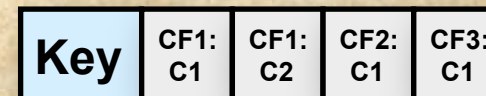
Source: 451 Research, used with permission

Major categories of NoSQL

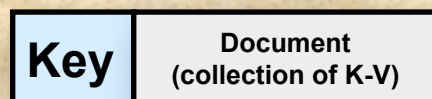
Key-Value store



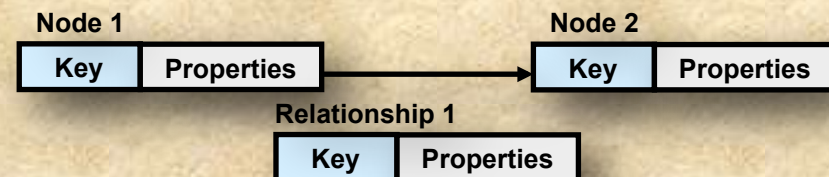
Column store

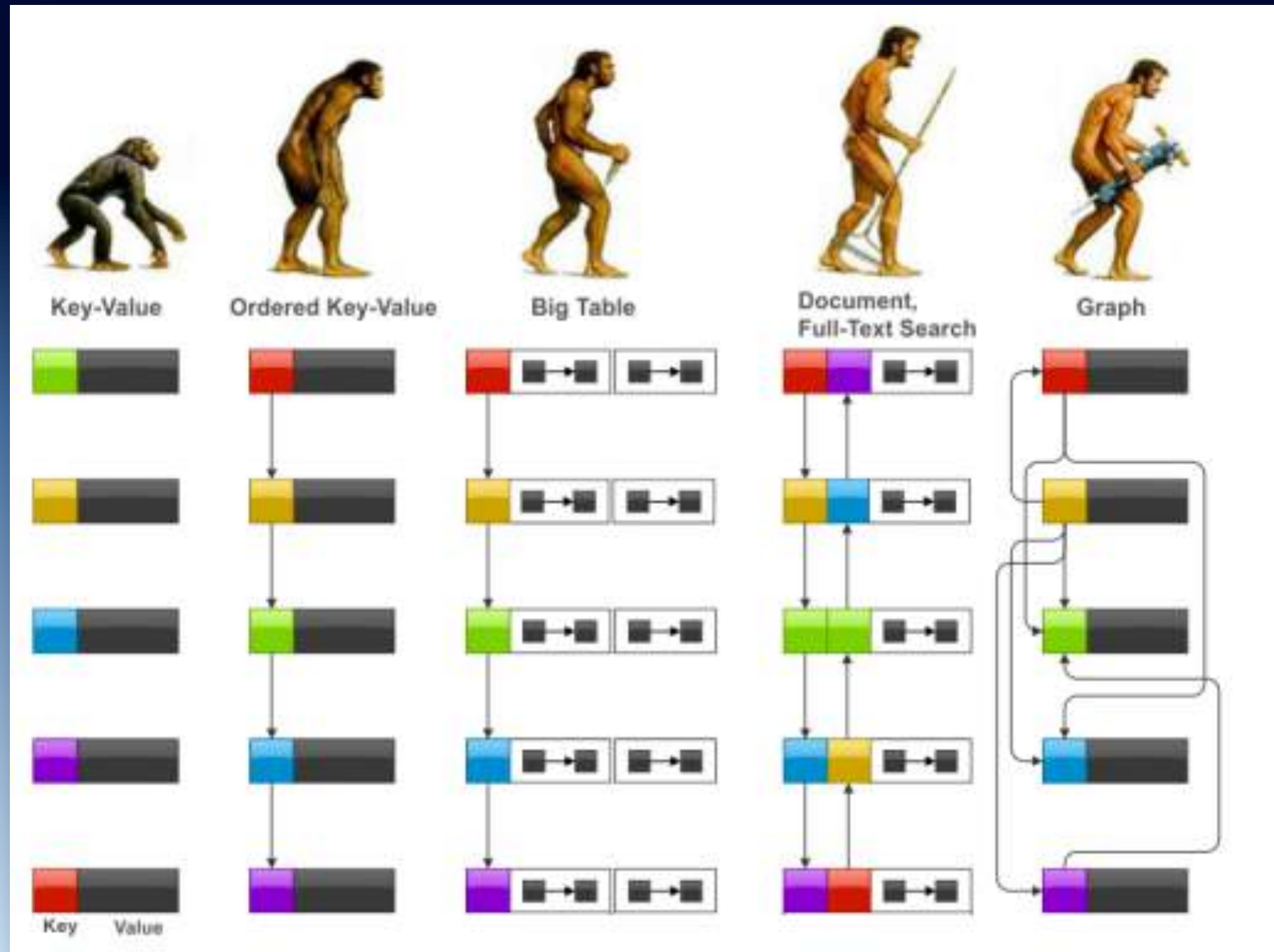


Document store








Graph store





Source: Ilya Katsov, used with permission

Popular NoSQL DBs

	License	Protocol	API/Query	Replication
	Apache	Thrift	CQL, Thrift	P2P
	Apache	REST/HTTP	JSON, MR	M-M
	AGPL	Proprietary	BSON	M-S, Shard
	BSD	Telnet-Like*	Many Langs.	M-S
	Apache	REST/HTTP	JSON, MR	P2P*





Source: "Big Data Projects: How to Choose NoSQL Databases" Thomas Casselberry (21 January 2015)

Analysis of replication consensus strategies

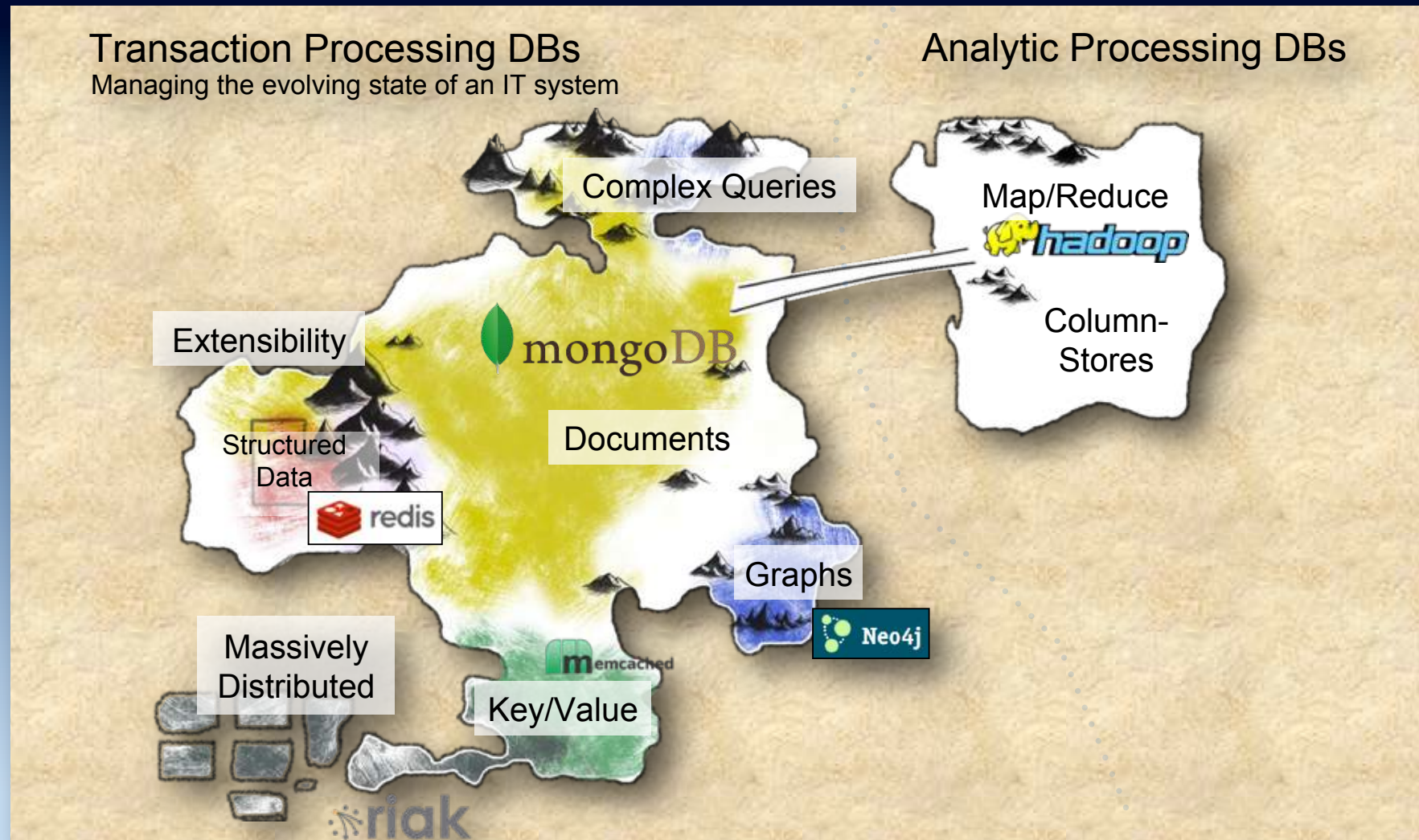
	Backups	M-S	M-M	2PC	Paxos
Consistency	Weak	Eventual		Strong	
Transactions	No	Full	Local	Full	
Latency	Low			High	
Throughput	High			Low	Medium
Data Loss	Lots	Some		None	
Failover	Down	R-only	R-W		

Source: "The Road to Akka Cluster and Beyond" Jonas Bonér (3 December 2013)

The rise of multi-model DBs ...

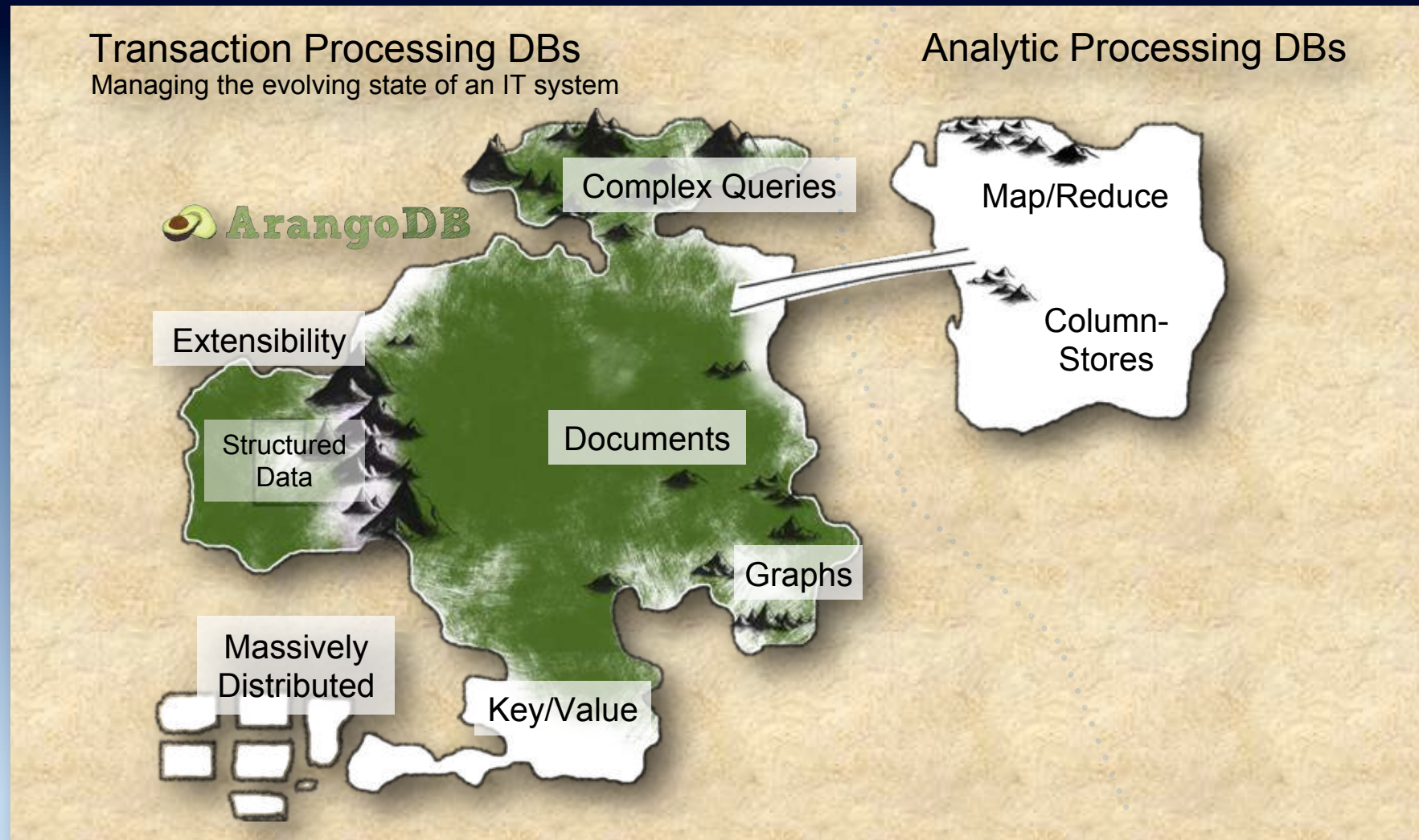
	K-V	Column	Document	Graph
 ArangoDB	✓		✓	✓
 Cassandra	✓	✓		✓*
	✓		✓	
 riak	✓		✓	

The rise of multi-model DBs ...



Source: ArangoDB, used with permission

The rise of multi-model DBs

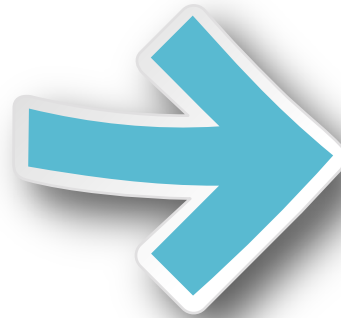


Source: ArangoDB, used with permission

Commercialization examples



Cassandra



basho

Key-Value store

- Simplest NoSQL stores, provide low-latency writes but single key/value access
- Store data as a hash table of keys where every key maps to an opaque binary object
- Easily scale across many machines
- Use-cases: applications that require massive amounts of simple data (sensor, web operations), applications that require rapidly changing data (stock quotes), caching

Redis and Riak examples

```
{
  database number: {
    "key 1": "value",
    "key 2": [ "value", "value",
              "value" ],
    "key 3": [
      { "value": "value", "score":
        score },
      { "value": "value", "score":
        score },
      ...
    ],
    "key 4": {
      "property 1": "value",
      "property 2": "value",
      "property 3": "value", ...
    }, ...
  }, ...
}
```

```
{
  "bucket 1": {
    "key 1": document + content-type,
    "key 2": document + content-type,
    "link to another object 1": URI of
    other bucket/key,
    "link to another object 2": URI of
    other bucket/key,
  },
  "bucket 2": {
    "key 3": document + content-type,
    "key 4": document + content-type,
    "key 5": document + content-type
    ...
  }, ...
}
```

DEMO



redis

Connection

```
Jedis j = new Jedis("localhost", 6379);  
j.connect();
```

```
System.out.println("Connected to Redis");
```

Create

```
String id = Long.toString(j.incr("global:nextUserId"));
```

```
j.set("uid:" + id + ":name", "akmal");
```

```
j.set("uid:" + id + ":age", "40");
```

```
j.set("uid:" + id + ":date", new Date().toString());
```

```
j.sadd("uid:" + id + ":likes", "satay");
```

```
j.sadd("uid:" + id + ":likes", "kebabs");
```

```
j.sadd("uid:" + id + ":likes", "fish-n-chips");
```

```
j.hset("uid:lookup:name", "akmal", id);
```

Read

```
String id = j.hget("uid:lookup:name", "akmal");  
  
print("name ", j.get("uid:" + id + ":name"));  
print("age ", j.get("uid:" + id + ":age"));  
print("date ", j.get("uid:" + id + ":date"));  
print("likes ", j.smembers("uid:" + id + ":likes"));
```

Update

```
String id = j.hget("uid:lookup:name", "akmal");
```

```
j.set("uid:" + id + ":age", "29");
```


Delete

```
String id = j.hget("uid:lookup:name", "akmal");
```

```
j.del("uid:" + id + ":name");
```

```
j.del("uid:" + id + ":age");
```

```
j.del("uid:" + id + ":date");
```

```
j.del("uid:" + id + ":likes");
```

Column store ...

- Manage structured data, with multiple-attribute access
- Columns are grouped together in “column-families/groups”; each storage block contains data from only one column/column set to provide data locality for “hot” columns
- Column groups defined *a priori*, but support variable schemas within a column group

Column store

- Scale using replication, multi-node distribution for high availability and easy failover
- Optimized for writes
- Use cases: high throughput verticals (activity feeds, message queues), caching, web operations

Cassandra example

```
{  
  "column family 1": {  
    "key 1": {  
      "property 1": "value",  
      "property 2": "value"  
    },  
    "key 2": {  
      "property 1": "value",  
      "property 4": "value",  
      "property 5": "value"  
    }  
  }, ...  
}
```

```
{  
  "column family 2": {  
    "super key 1": {  
      "key 1": {  
        "property 1": "value",  
        "property 2": "value"  
      },  
      "key 2": {  
        "property 1": "value",  
        "property 4": "value",  
        "property 5": "value"  
      }, ...  
    }, ...  
  }, ...  
}
```

DEMO



cassandra

Connection

```
Class.forName("org.apache.cassandra.cql.jdbc.CassandraDriver");  
connection = DriverManager.getConnection(  
    "jdbc:cassandra://localhost:9160/demodb");  
  
System.out.println("Connected to Cassandra");
```

Create

```
String query =  
"BEGIN BATCH\n" +  
"INSERT INTO people (name, age, date, likes) VALUES ('akmal', 40, ""  
+ new Date() +  
"", {'satay', 'kebabs', 'fish-n-chips'})\n" +  
"APPLY BATCH;";
```

```
Statement statement = connection.createStatement();  
statement.executeUpdate(query);  
statement.close();
```

Read

```
String query = "SELECT * FROM people";
```

```
Statement statement = connection.createStatement();
```

```
ResultSet cursor = statement.executeQuery(query);
```

```
while (cursor.next())
```

```
    for (int j = 1; j < cursor.getMetaData().getColumnCount()+1; j++)
```

```
        System.out.printf("%-10s: %s%n",
```

```
            cursor.getMetaData().getColumnName(j),
```

```
            cursor.getString(cursor.getMetaData().getColumnName(j)));
```

```
cursor.close();
```

```
statement.close();
```


Update

```
String query =  
    "UPDATE people SET age = 29 WHERE name = 'akmal';"
```

```
Statement statement = connection.createStatement();  
statement.executeUpdate(query);  
statement.close();
```

Delete

```
String query =  
"BEGIN BATCH\n" +  
"DELETE FROM people WHERE name = 'akmal'\n" +  
"APPLY BATCH;";
```

```
Statement statement = connection.createStatement();  
statement.executeUpdate(query);  
statement.close();
```

Document store

- Represent rich, hierarchical data structures, reducing the need for multi-table joins
- Structure of the documents need not be known *a priori*, can be variable, and evolve instantly, but a query can understand the contents of a document
- Use cases: rapid ingest and delivery for evolving schemas and web-based objects

MongoDB example

```
{  
  "namespace 1": any json object,  
  "namespace 2": any json object,  
  ...  
}
```

```
{  
  "namespace 1": [  
    {  
      "_id": "key 1",  
      "property 1": "value",  
      "property 2": {  
        "property 3": "value",  
        "property 4": [ "value",  
          "value", "value" ]  
      }, ...  
    },  
    ...  
  ]  
}
```

DEMO



Connection

```
private static final String DBNAME = "demodb";  
private static final String COLLNAME = "people";  
  
...  
MongoClient mongoClient = new MongoClient("localhost", 27017);  
DB db = mongoClient.getDB(DBNAME);  
DBCollection collection = db.getCollection(COLLNAME);  
  
System.out.println("Connected to MongoDB");
```

Create

```
BasicDBObject document = new BasicDBObject();
```

```
List<String> likes = new ArrayList<String>();
```

```
likes.add("satay");
```

```
likes.add("kebabs");
```

```
likes.add("fish-n-chips");
```

```
document.put("name", "akmal");
```

```
document.put("age", 40);
```

```
document.put("date", new Date());
```

```
document.put("likes", likes);
```

```
collection.insert(document);
```

Read

```
BasicDBObject document = new BasicDBObject();  
document.put("name", "akmal");
```

```
DBCursor cursor = collection.find(document);
```

```
while (cursor.hasNext())  
    System.out.println(cursor.next());
```

```
cursor.close();
```


Update

```
BasicDBObject document = new BasicDBObject();  
document.put("name", "akmal");
```

```
BasicDBObject newDocument = new BasicDBObject();  
newDocument.put("age", 29);
```

```
BasicDBObject updateObj = new BasicDBObject();  
updateObj.put("$set", newDocument);
```

```
collection.update(document, updateObj);
```

Delete

```
BasicDBObject document = new BasicDBObject();  
document.put("name", "akmal");  
  
collection.remove(document);
```

DEMO



mongoDB



TM

Connection

```
var async = require('async');  
var MongoClient = require('mongodb').MongoClient;  
MongoClient.connect("mongodb://localhost:27017/demodb",  
function(err, db) {  
  if (err) {  
    return console.log(err);  
  }  
  console.log("Connected to MongoDB");  
  var collection = db.collection('people');  
  var document = {  
    'name':'akmal',  
    'age':40,  
    'date':new Date(),  
    'likes':['satay', 'kebabs', 'fish-n-chips']  
  };  
};
```

Create

```
function (callback) {  
  collection.insert(document, {w:1}, function(err, result) {  
    if (err) {  
      return callback(err);  
    }  
    callback();  
  });  
},
```

Read

```
function (callback) {  
  collection.findOne({'name':'akmal'}, function(err, item) {  
    if (err) {  
      return callback(err);  
    }  
    console.log(item);  
    callback();  
  });  
},
```

Update

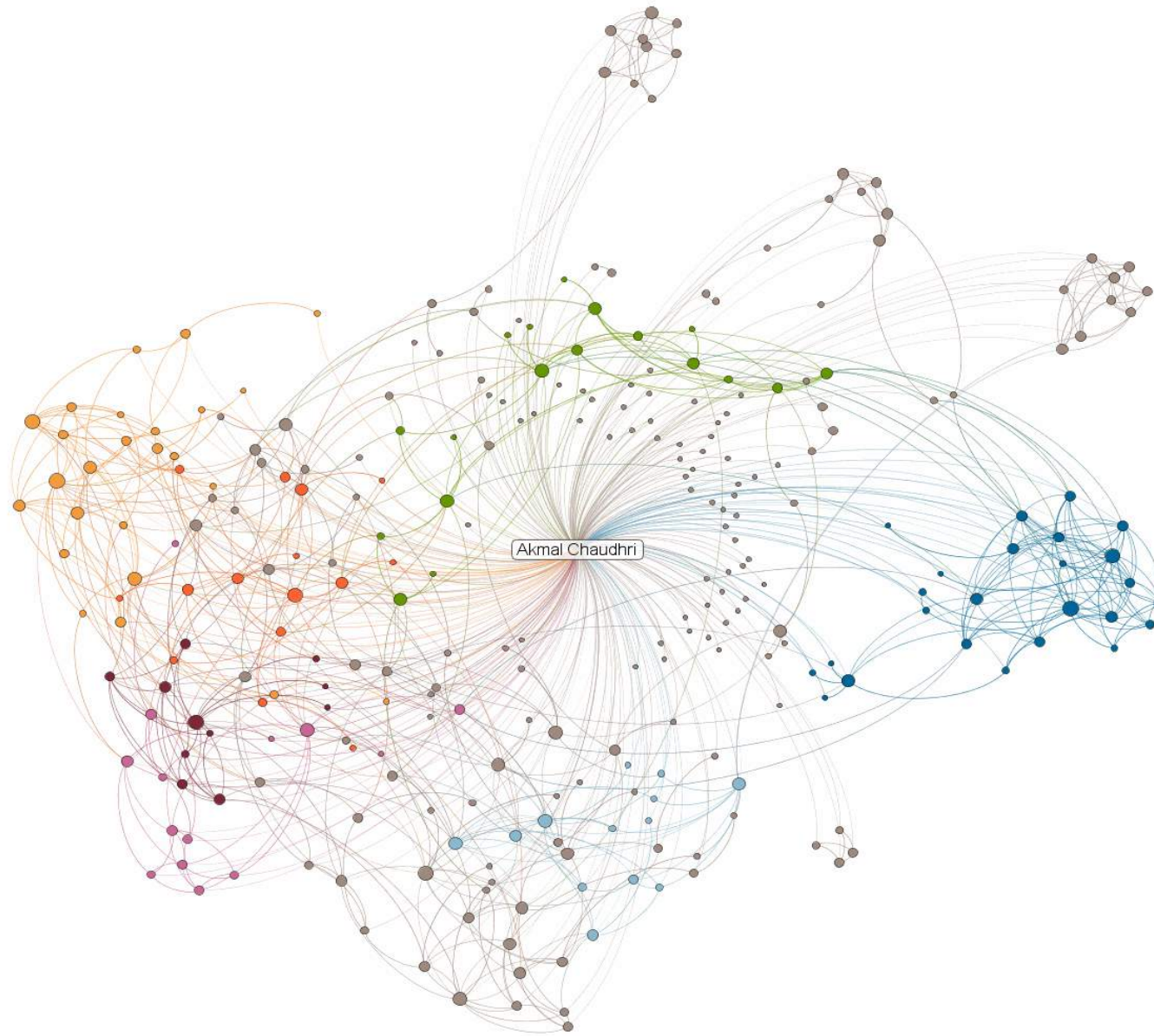
```
function (callback) {  
  collection.update({'name':'akmal'}, {$set: {'age':29}}, {w:1},  
  function(err, result) {  
    if (err) {  
      return callback(err);  
    }  
    callback();  
  });  
},
```

Delete

```
function (callback) {  
  collection.remove({'name':'akmal'}, function(err, result) {  
    if (err) {  
      return callback(err);  
    }  
    callback();  
  });  
},
```


Graph store

- Use nodes, relationships between nodes, and key-value properties
- Access data using graph traversal, navigating from start nodes to related nodes according to graph algorithms
- Faster for associative data sets
- Use cases: storing and reasoning on complex and connected data, such as inferencing applications in healthcare, government, telecom, oil, performing closure on social networking graphs



DEMO



Neo4j
the graph database

Connection

```
private static final String DB_PATH =  
    "C:/neo4j-community-1.8.2/data/graph.db";  
  
private static enum RelTypes implements RelationshipType {  
    LIKES  
}  
...  
graphDb =  
    new GraphDatabaseFactory().newEmbeddedDatabase(DB_PATH);  
registerShutdownHook(graphDb);  
  
System.out.println("Connected to Neo4j");
```

Create

```
Transaction tx = graphDb.beginTx();
```

```
try {  
    firstNode = graphDb.createNode();  
    firstNode.setProperty("name", "akmal");  
    firstNode.setProperty("age", 40);  
    firstNode.setProperty("date", new Date().toString());  
    secondNode = graphDb.createNode();  
    secondNode.setProperty("food", "satay, kebabs, fish-n-chips");  
    relationship = firstNode.createRelationshipTo(secondNode,  
        RelTypes.LIKES);  
    relationship.setProperty("likes", "likes");  
    tx.success();  
} finally { tx.finish(); }
```

Read

```
Transaction tx = graphDb.beginTx();
```

```
try {  
    print("name", firstNode.getProperty("name"));  
    print("age", firstNode.getProperty("age"));  
    print("date", firstNode.getProperty("date"));  
    print("likes", secondNode.getProperty("food"));  
    tx.success();  
} finally { tx.finish(); }
```

Update

```
Transaction tx = graphDb.beginTx();
```

```
try {  
    firstNode.setProperty("age", 29);  
    tx.success();  
} finally { tx.finish(); }
```

Delete

```
Transaction tx = graphDb.beginTx();
```

```
try {  
    firstNode.getSingleRelationship(RelTypes.LIKES,  
        Direction.OUTGOING).delete();  
    firstNode.delete();  
    secondNode.delete();  
    tx.success();  
} finally { tx.finish(); }
```


NoSQL use cases ...

- Online/mobile gaming
 - Leaderboard (high score table) management
 - Dynamic placement of visual elements
 - Game object management
 - Persisting game/user state information
 - Persisting user generated data (e.g. drawings)
- Display advertising on web sites
 - Ad Serving: match content with profile and present
 - Real-time bidding: match cookie profile with advert inventory, obtain bids, and present advert

NoSQL use cases

- Dynamic content management and publishing (news and media)
 - Store content from distributed authors, with fast retrieval and placement
 - Manage changing layouts and user generated content
- E-commerce/social commerce
 - Storing frequently changing product catalogs
- Social networking/online communities
- Communications
 - Device provisioning

Use case requirements ...

- Schema flexibility and development agility
 - Application not constrained by fixed pre-defined schema
 - Application drives the schema
 - Ability to develop a minimal application rapidly, and iterate quickly in response to customer feedback
 - Ability to quickly add, change or delete “fields” or data-elements
 - Ability to handle mix of structured, unstructured data
 - Easier, faster programming, so faster time to market and quick to adapt

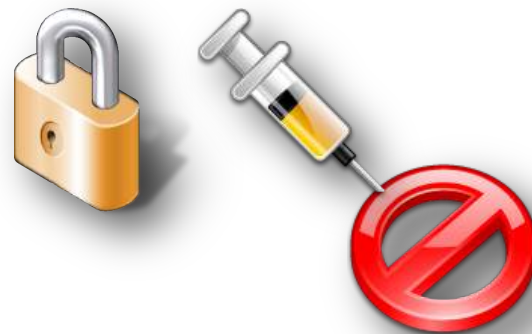
Use case requirements ...

- Consistent low latency, even under high load
 - Typically milliseconds or sub-milliseconds, for reads and writes
 - Even with millions of users
- Dynamic elasticity
 - Rapid horizontal scalability
 - Ability to add or delete nodes dynamically
 - Application transparent elasticity, such as automatic (re)distribution of data, if needed
 - Cloud compatibility

Use case requirements

- High availability
 - 24 x 7 x 365 availability
 - (Today) Requires data distribution and replication
 - Ability to upgrade hardware or software without any down time
- Low cost
 - Commonly available hardware
 - Lower cost software, such as open source or pay-per-use in cloud
 - Reduced need for database admin and maintenance

Security and vulnerability



Security



Source: Shutterstock Image ID 134699780

NoSQL databases threat model

1. Transactional integrity
2. Lax authentication mechanisms
3. Inefficient authorization mechanisms
4. Susceptibility to injection attacks
5. Lack of consistency
6. Insider attacks

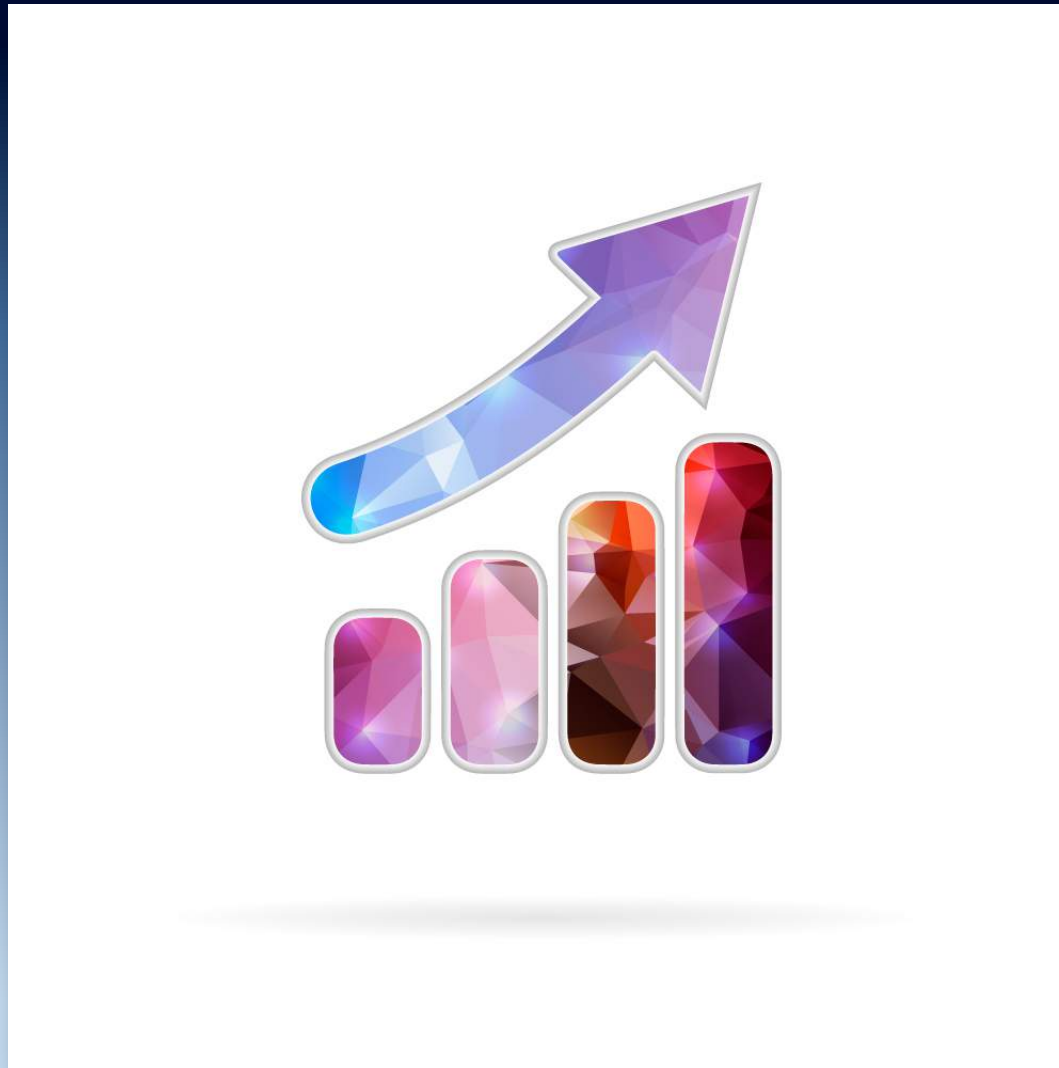
NoSQL data security issues

1. Data at rest
2. Data in motion (client-node communications)
3. Data in motion (inter-node communications)
4. Authentication
5. Authorization
6. Audit
7. Data consistency
8. NoSQL injection exploits

5 Big Data security pitfalls

1. Running databases in a “trusted” environment
2. Loose access control
3. Static protection schemes
4. Inadequate solutions for detecting sensitive data
5. Lack of entitlement, auditing and monitoring

Security problems increasing



Source: Shutterstock Image ID 216333160

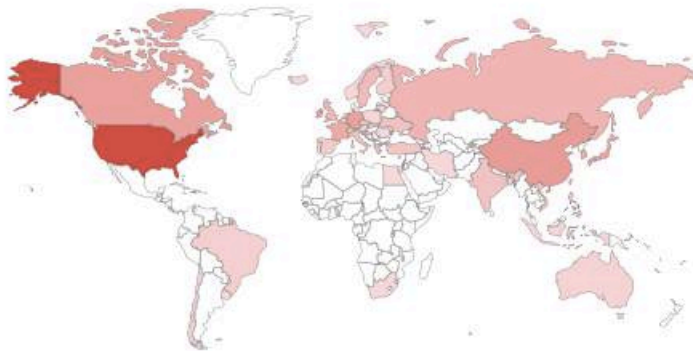
Well-known ports

Product	Ports
MongoDB	27017, 28017, 27080
CouchDB	5984
HBase	9000
Cassandra	9160
Neo4j	7474
Redis	6379
Riak	8098

Shodan port example

Cassandra NoSQL Database

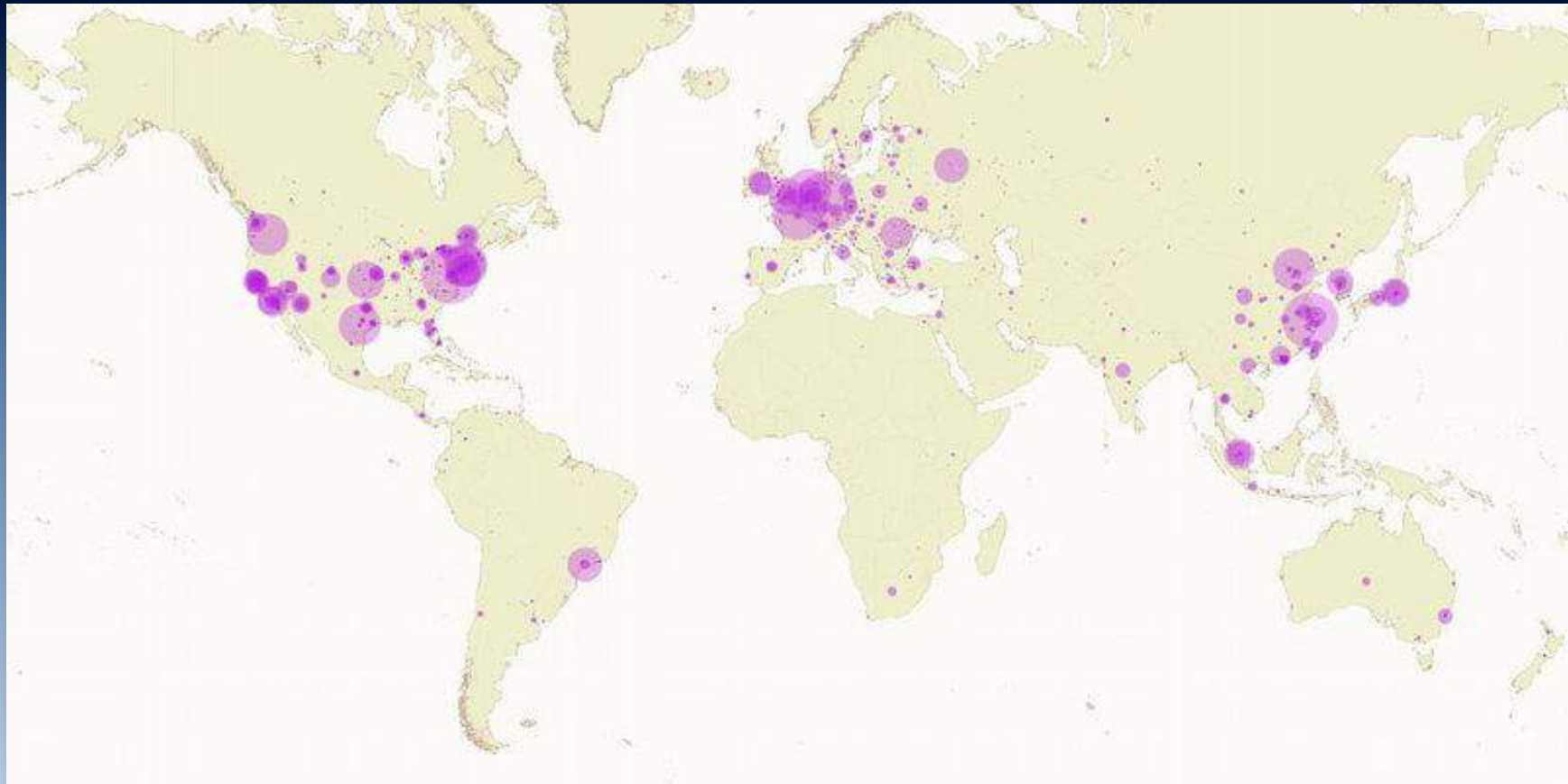
Search for `port:9160` returned 3,311 results on 10-10-2014



Top Countries

1. United States	2,361
2. China	126
3. Canada	92
4. Germany	89
5. Ireland	88
6. Japan	73
7. France	53
8. Korea, Republic of	41
9. Singapore	39
10. Russian Federation	27

~40,000 MongoDB open online



Source: “MongoDB databases at risk” Jens Heyens, Kai Greshake and Eric Petryka (January 2015)

MongoDB leaking data

Product	Instances	Size (TB)
MongoDB	29,980	595.2



Source: "It's the Data, Stupid!" John Matherly (18 July 2015)

NoSQL apps leaking data ...

Product	Instances	Size (TB)
Redis	35,330	13.21-17.08
MongoDB	39,134	619.80
Memcached	118,574	11.35
ElasticSearch	8990	531.20



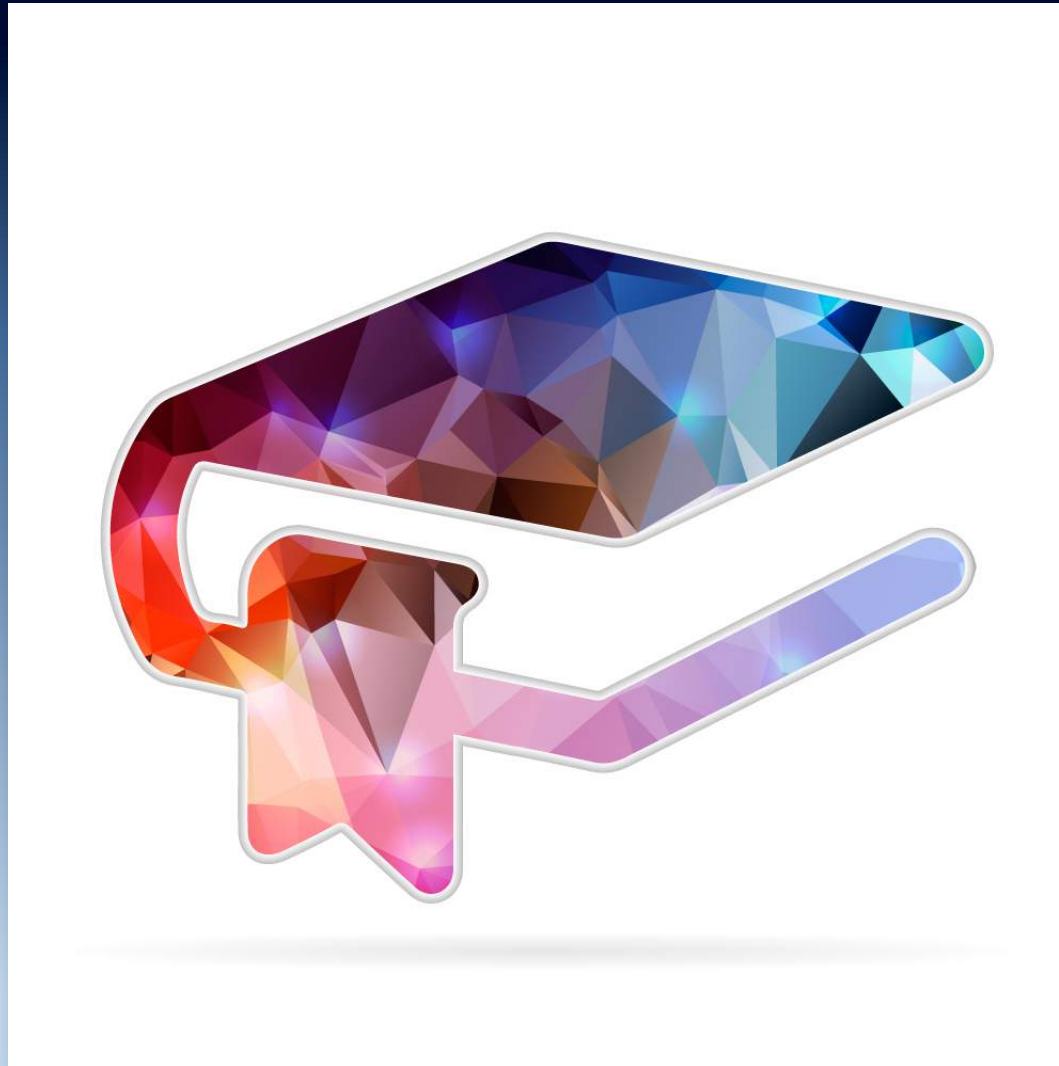
Source: "Data, Technologies and Security - Part 1" BinaryEdge (14 August 2015)

NoSQL apps leaking data



These technologies' default settings tend to have no configuration for authentication, encryption, authorization or any other type of security controls that we take for granted. Some of them don't even have a built-in access control.

Read the manual



Source: Shutterstock Image ID 196307192

Redis security



Redis is designed to be accessed by trusted clients inside trusted environments. This means that usually it is not a good idea to expose the Redis instance directly to the internet or, in general, to an environment where untrusted clients can directly access the Redis TCP port or UNIX socket.

MongoDB security



The most effective way to reduce risk for MongoDB deployments is to run your entire MongoDB deployment, including all MongoDB components (i.e. mongod, mongos and application instances) in a trusted environment.

Memcached security



Memcached has no security or authentication. Please ensure that your server is appropriately firewalled, and that the port(s) used for memcached servers are not publicly accessible. Otherwise, anyone on the internet can put data into and read data from your cache.

CouchDB security



*When you start out fresh, CouchDB allows any request to be made by anyone ... While it is incredibly easy to get started with CouchDB that way, it should be obvious that putting a default installation into the wild is adventurous. Any rogue client could come along and delete a database. **relax***

NoSQL injection attacks ...

- NoSQL systems are vulnerable
- Various types of attacks
- Understand the vulnerabilities and consequences



NoSQL injection attacks

- Popular NoSQL products will attract more interest and scrutiny
- Features of some programming languages, e.g. PHP
- Server-Side JavaScript (SSJS)



NoSQL injection testing

- NoSQLMap project
 - Open source proof-of-concept Python tool
 - Automates injection attacks
 - Exploits MongoDB vulnerabilities
 - Future support for other NoSQL databases

Polyglot persistence



Source: Heroku, used with permission

Polyglot persistence



Source: Adapted from "PolyglotPersistence" Martin Fowler (16 November 2011)

But ...



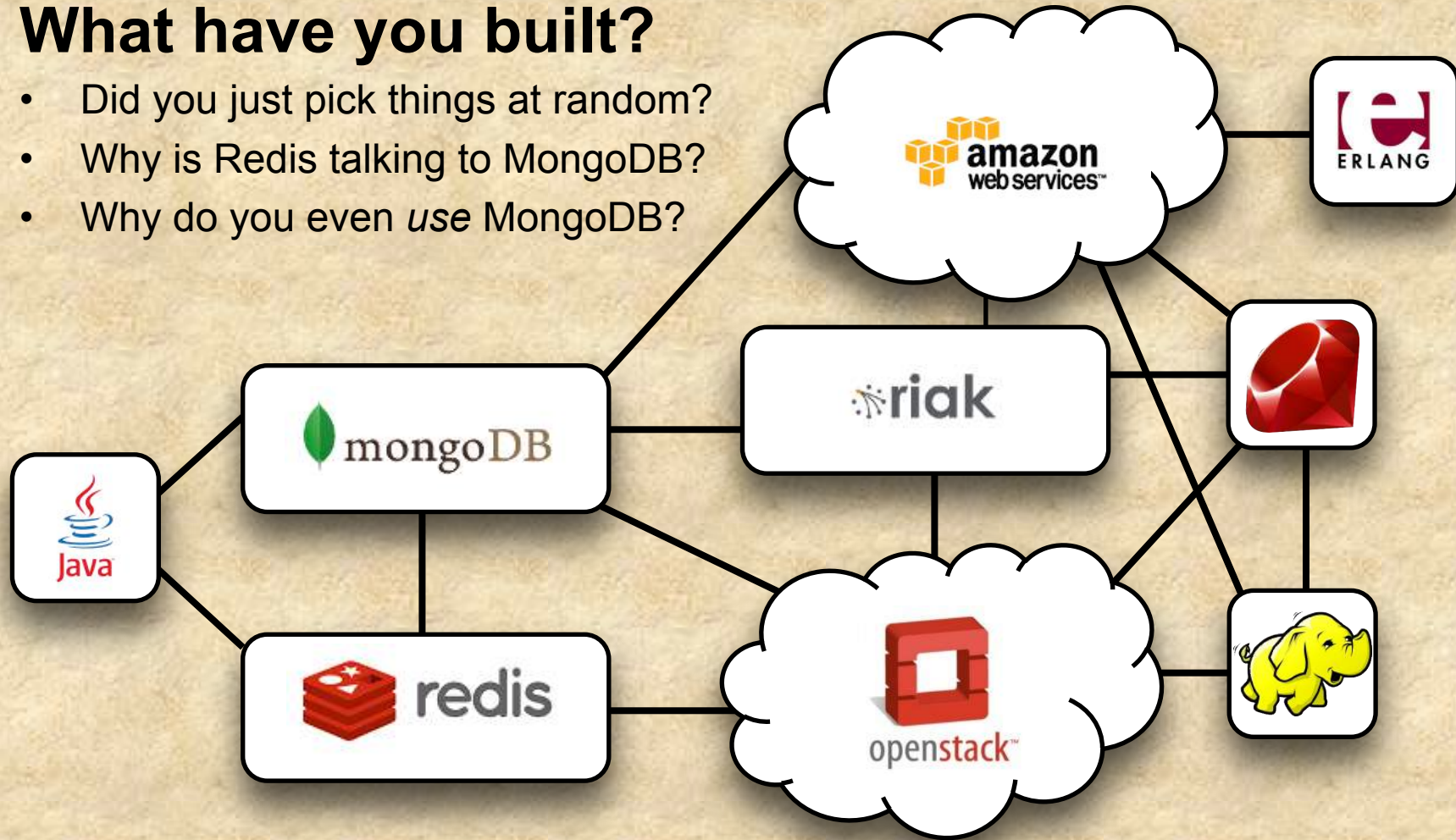
In an often-cited post on polyglot persistence, Martin Fowler sketches a web application for a hypothetical retailer that uses each of Riak, Neo4j, MongoDB, Cassandra, and an RDBMS for distinct data sets. It's not hard to imagine his retailer's DevOps engineers quitting in droves.

-- Stephen Pimentel

And ...

What have you built?

- Did you just pick things at random?
- Why is Redis talking to MongoDB?
- Why do you even *use* MongoDB?



Source: After <https://twitter.com/codinghorror/status/347070841059692545/>

Polyglot persistence ...

- Multiple developer skills
 - The programmer must learn new languages and APIs
- Multiple DBA skills
 - The DBA must learn new backup/recovery utilities and new optimization techniques
- Multiple analyst skills
 - The analyst must study new database concepts and how to model them best

Polyglot persistence ...



What I've seen in the past has been is if you try to take on six of these [technologies], you need a staff of 18 people minimum just to operate the storage side - say, six storage technologies. That's not scalable and it's too expensive.

-- Dave McCrory

Polyglot persistence

- Different APIs
 - Develop public API for each NoSQL store (Disney)

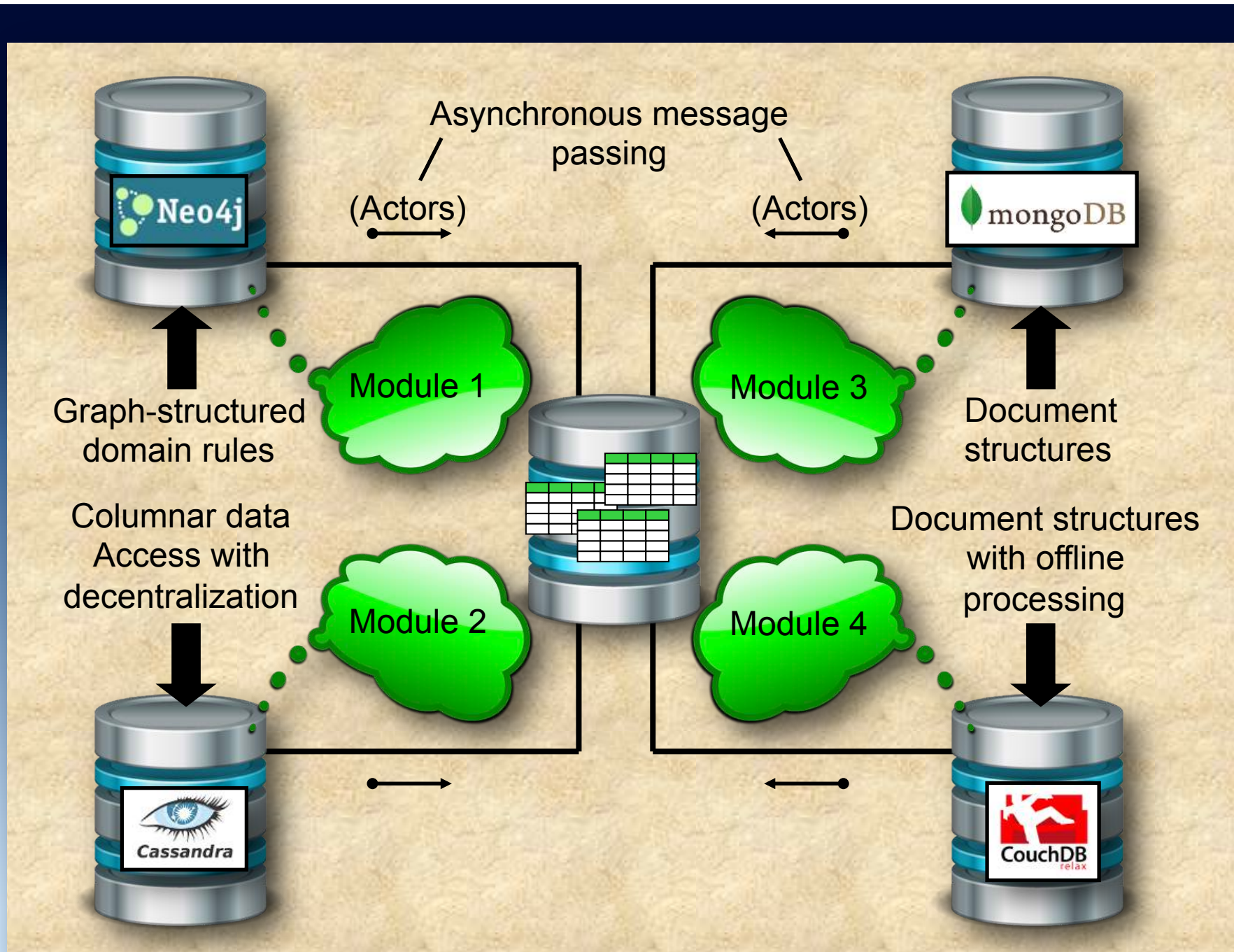
Public API for NoSQL store



In some cases, the team decided to hide the platform's complexity from users; not to facilitate its use, but to keep loose-cannon developers from doing something crazy that could take down the whole cluster. It could show them all the controls and knobs in a NoSQL database, but "they tend to shoot each other," Jacob said. "First they shoot themselves, then they shoot each other."

Polyglot persistence examples

- Disney
 - Cassandra, Hadoop, MongoDB
- Interactive Mediums
 - CouchDB, MySQL
- Mendeley
 - HBase, MongoDB, Solr, Voldemort
- Netflix
 - Cassandra, Hadoop/HBase, RDBMS, SimpleDB
- Twitter
 - Cassandra, FlockDB, Hadoop/HBase, MySQL



Source: Debasish Ghosh, used with permission

Multi-paradigm example

- Application that routes picking baskets for inventory in a warehouse
- A graph with bins of inventory (nodes) along aisles (edges)
- Store graph in Neo4j for performance
- Asynchronously persist in MySQL for reporting
- Move data using asynchronous message queue
- Faster performance, easier development, simpler scaling, and reduced cost

Polyglot persistence with EclipseLink JPA

- Java Persistence API (JPA) for access to NoSQL systems
- Annotations and XML to identify stored NoSQL entities
- An application can use multiple database systems
- Single composite Persistence Unit (PU) supports relational and non-relational data
- Support for MongoDB and Oracle NoSQL with other products planned

Benchmarks and performance

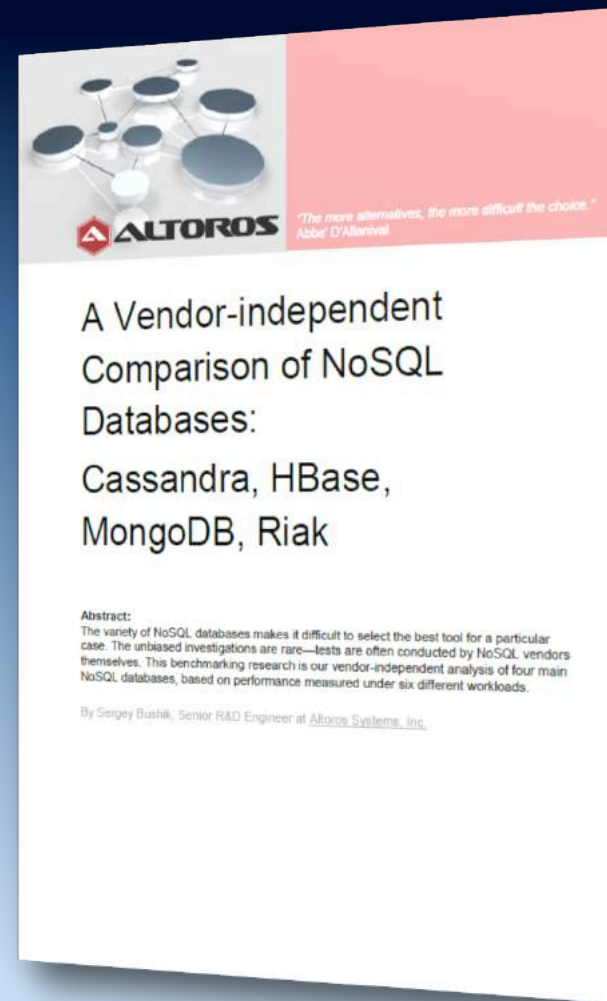


Yahoo Cloud Serving BM ...

- Originally Tested Systems
 - Cassandra, HBase, Yahoo!'s PNUTS, sharded MySQL
- Tier 1 (performance)
 - Latency by increasing the server load
- Tier 2 (scalability)
 - Scalability by increasing the number of servers

Yahoo Cloud Serving BM

- Yahoo Cloud Serving Benchmark (YCSB)
 - Research paper
 - Slide deck
- Various reports
 - See resources



2015 YCSB results ...



Comparing Couchbase Server 3.0.2 with MongoDB 3.0.2 Benchmark Results and Analysis

Composed by Avalon Consulting, LLC

Big thinkers for Big Data

We rate emerging technologies to give
you a competitive advantage.



Introduction

The data needs of today's Enterprise require a special set of tools. At the center of these tools lies the NoSQL database. In recent years, NoSQL has become a growing part of the modern Enterprise infrastructure. Knowing how to implement a highly scalable NoSQL database that fits current and future use cases and scales easily and efficiently is critical in satisfying these ever-growing demands.

It's important to consider performance, scalability, consistency, and availability when selecting a NoSQL database. However, this benchmark focuses exclusively on performance. In an era where applications may have to support millions of users and where users expect faster and faster responses, performance can be the deciding factor between success and failure. A high



HIGH PERFORMANCE BENCHMARKING: MongoDB and NoSQL Systems



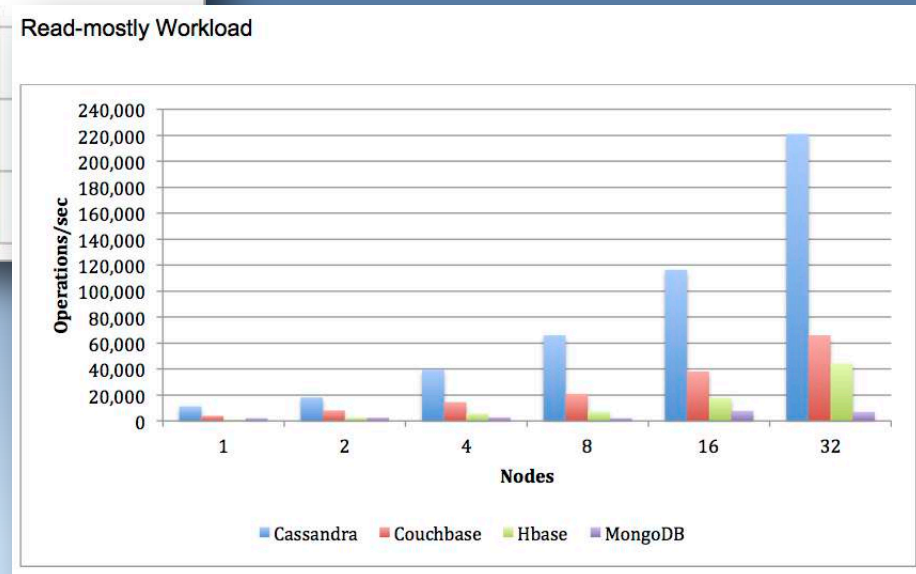
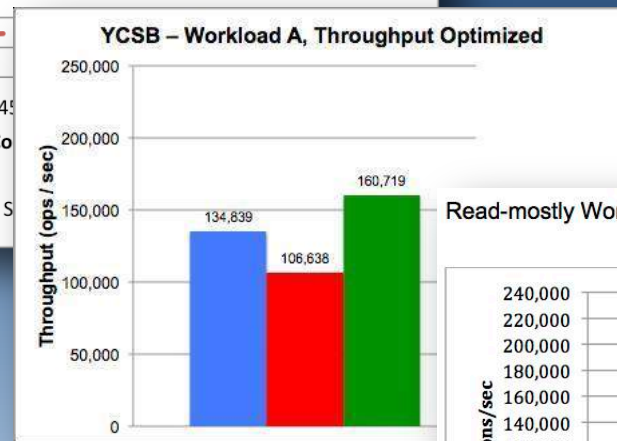
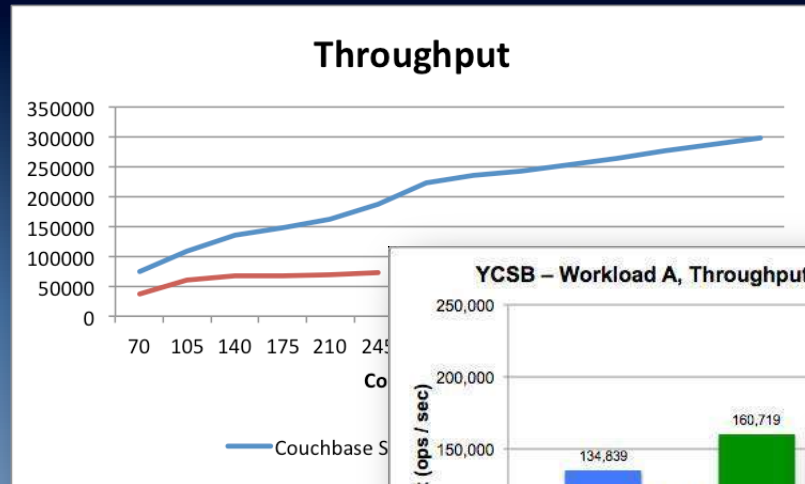
Benchmarking Top NoSQL Databases

Apache Cassandra, Couchbase, HBase, and MongoDB

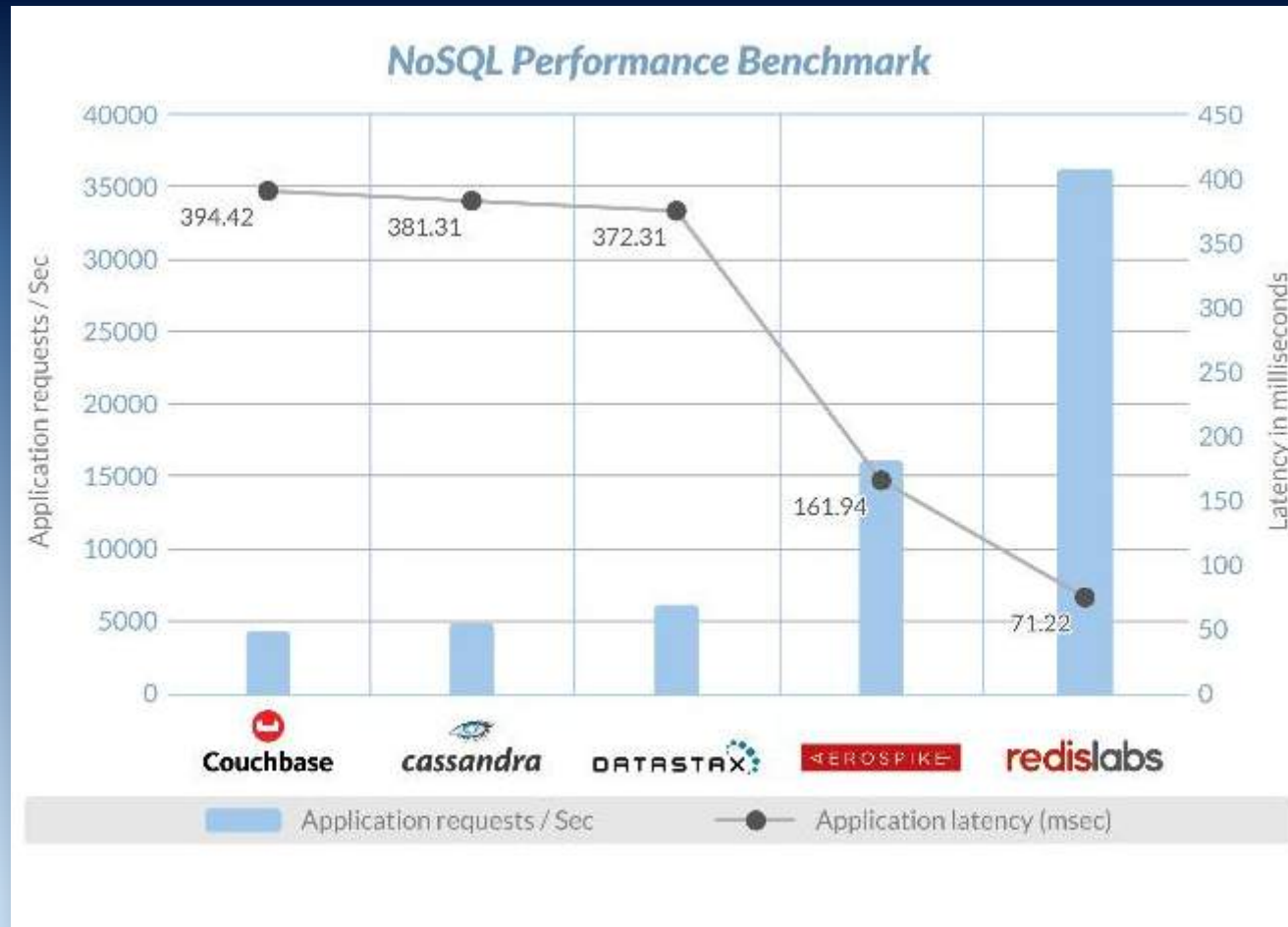
April 13, 2015

<http://www.endpoint.com/>

2015 YCSB results

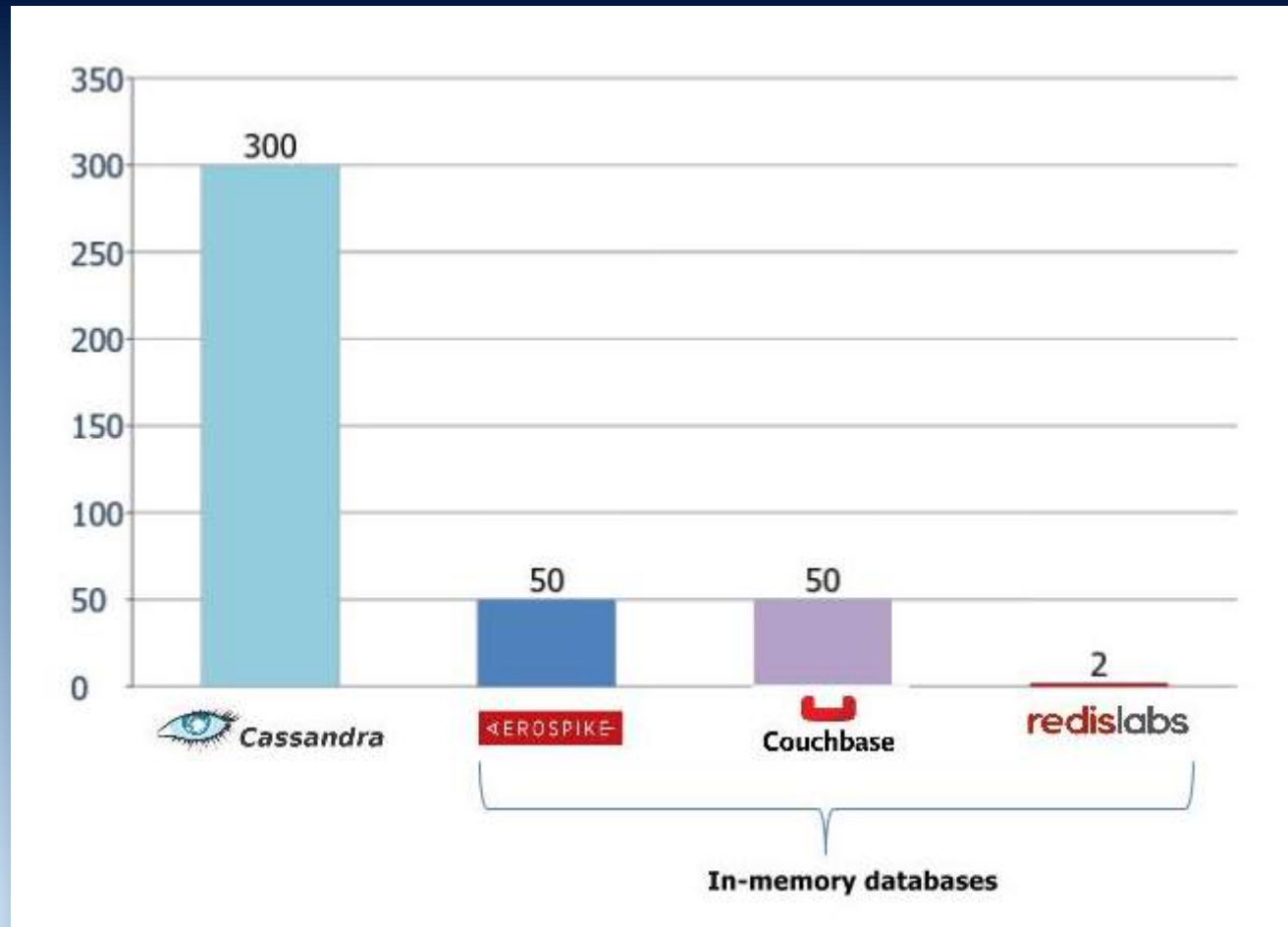


Redis customer benchmark



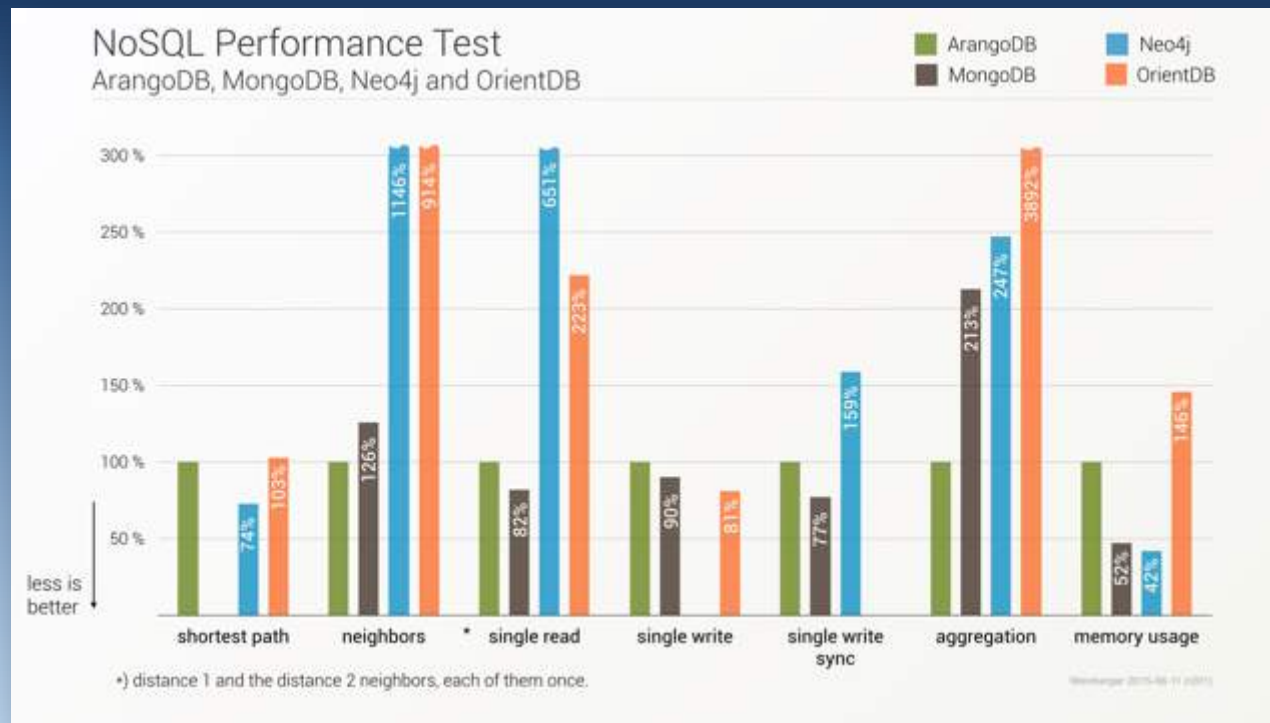
Source: "Busting 4 Myths About In-Memory Databases" Yiftach Shoolman (16 September 2015)

How many servers to get 1 million writes/sec on GCE?



Source: "Busting 4 Myths About In-Memory Databases" Yiftach Shoolman (16 September 2015)

Multi-model benchmark



Source: “How an open-source competitive benchmark helped to improve databases” Frank Celler (25 June 2015)

But ...



... any person who designs a benchmark is in a 'no win' situation, i.e. he can only be criticized. External observers will find fault with the benchmark as artificial or incomplete in one way or another. Vendors who do poorly on the benchmark will criticize it unmercifully.

-- Mike Stonebraker

“Can the Elephants Handle the NoSQL Onslaught?”

- DSS Workload (TPC-H)
 - Hive vs. Parallel Data Warehouse
- Modern OLTP Workload (YCSB)
 - MongoDB vs. SQL Server
- Conclusions
 - NoSQL systems are behind relational systems in performance

Linked Data Benchmark Council



- EU-funded project
- Develop Graph and RDF benchmarks

Jepsen stress testing ...

- Jepsen project
 - Rigorously test how various database systems handle partitions
 - Evaluate consistency
- Conclusions
 - Don't rely on vendor marketing, product documentation or "pull the plug" test

Jepsen stress testing

- Postgres
- Redis
- MongoDB
- Riak
- Zookeeper
- NuoDB
- Kafka
- Cassandra
- Redis redux
- RabbitMQ
- etcd and Consul
- Elasticsearch
- MongoDB stale reads
- Elasticsearch 1.5.0
- Aerospike
- Chronos
- MariaDB Galera Cluster

SSDs and log-structured I/O

- Database systems that use log-structured I/O have interference effects with SSDs that slow performance and increase latency
- The log-structured Flash Translation Layer (FTL) that makes flash look like a disk adversely interacts with the already log-structured I/O from the application

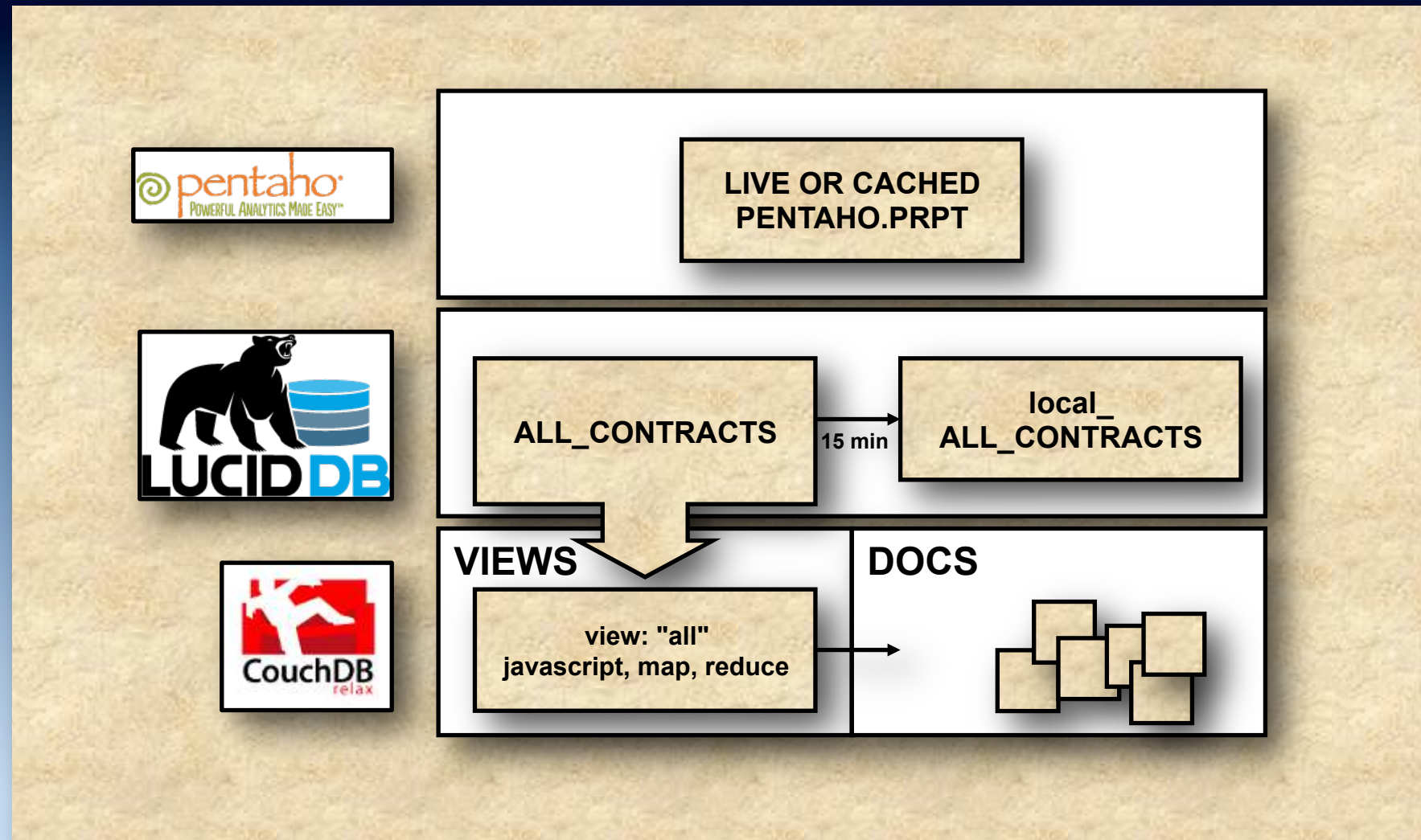
BI/Analytics



Architectures

- NoSQL reports
- NoSQL thru and thru
- NoSQL + MySQL
- NoSQL as ETL source
- NoSQL programs in BI tools
- NoSQL via BI database (SQL)

NoSQL via BI database (SQL)



Source: "SQL access to CouchDB views : Easy Reporting" Nicholas Goodman (22 June 2011)

DEMO



mongoDB



pentaho
POWERFUL ANALYTICS MADE EASY™

NoSQL alternatives

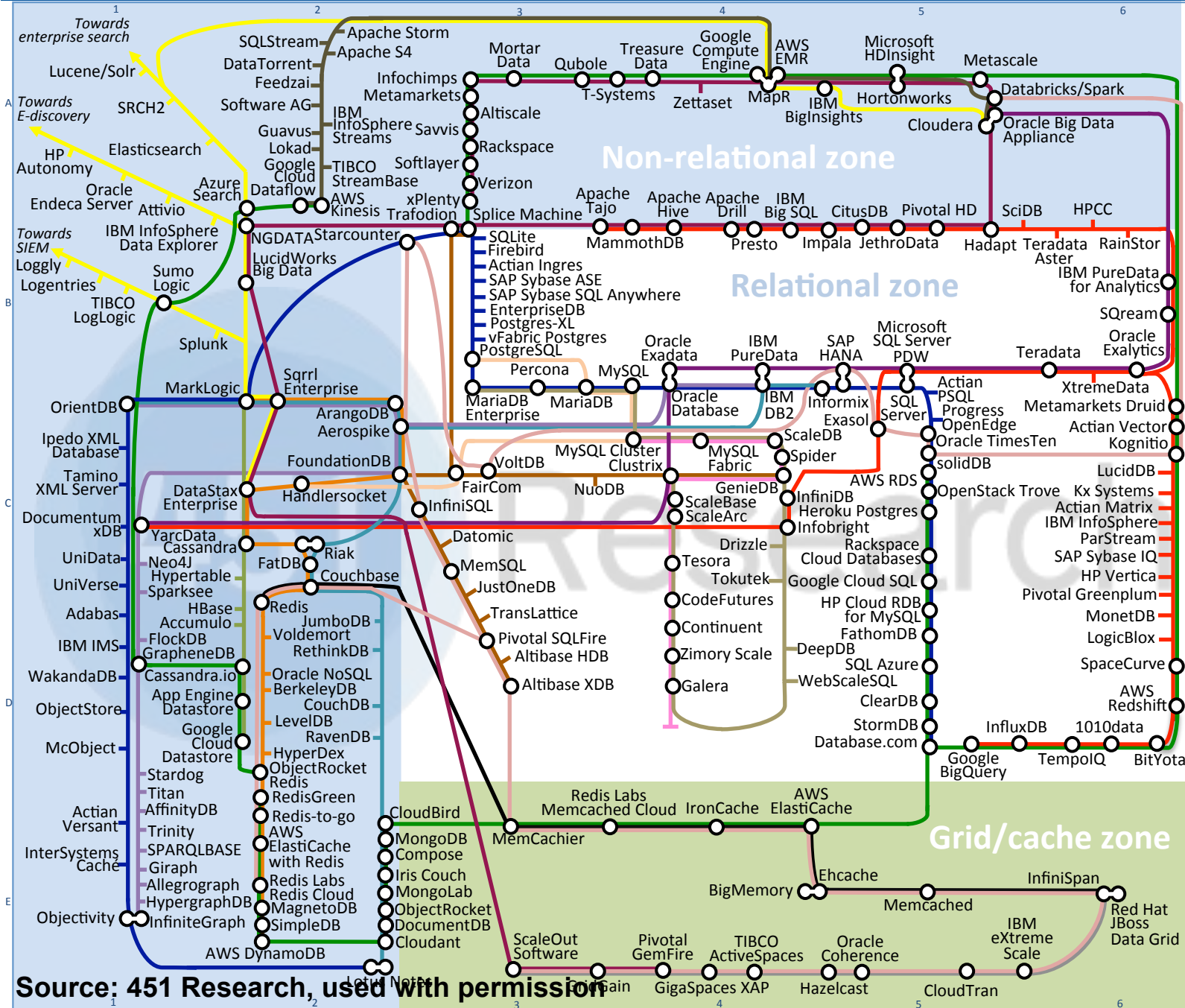


451 Research: Data Platforms Landscape Map – September 2014



© 2014 by 451 Research LLC. All rights reserved

- Key:**
- General purpose
 - Specialist analytic
 - -as-a-Service
 - BigTables
 - Graph
 - Document
 - Key value stores
 - Key value direct access
 - Hadoop
 - MySQL ecosystem
 - Advanced clustering/sharding
 - New SQL databases
 - Data caching
 - Data grid
 - Search
 - Appliances
 - In-memory
 - Stream processing



Source: 451 Research, used with permission

NewSQL

- Today, new challenges and requirements
 - “Web changes everything”
- Need more OLTP throughput
- Need real-time analytics
- ACID support
- Preserve SQL
 - Automatic query optimization
- Preserve investment
 - Existing skills and tools

DEMO



NUODB

Connection

```
Class.forName("com.nuodb.jdbc.Driver");
```

```
Properties properties = new Properties();
```

```
properties.put("user", "dba");
```

```
properties.put("password", "goalie");
```

```
properties.put("schema", "test");
```

```
connection = DriverManager.getConnection(  
    "jdbc:com.nuodb://localhost/test", properties);
```

```
System.out.println("Connected to NuoDB");
```

Create

```
PreparedStatement statement = connection.prepareStatement(  
    "INSERT INTO people (name, age, date, likes) VALUES (?, ?, ?, ?)");
```

```
statement.setString(1, "akmal");  
statement.setInt(2, 40);  
statement.setString(3, new Date().toString());  
statement.setString(4, "satay kebabs fish-n-chips");  
statement.addBatch();  
statement.executeBatch();  
connection.commit();
```

Read

```
String query = "SELECT * FROM people;";
```

```
Statement statement = connection.createStatement();
```

```
ResultSet cursor = statement.executeQuery(query);
```

```
while (cursor.next()) {
```

```
    System.out.print(cursor.getString(1) + " ");
```

```
    System.out.print(cursor.getInt(2) + " ");
```

```
    System.out.print(cursor.getString(3) + " ");
```

```
    System.out.println(cursor.getString(4));
```

```
}
```

```
cursor.close();
```

```
statement.close();
```

Update

```
String query =
```

```
"UPDATE people SET age = 29 WHERE name = 'akmal';";
```

```
Statement statement = connection.createStatement();
```

```
statement.executeUpdate(query);
```

```
connection.commit();
```

```
readData(connection);
```

Delete

```
String query = "DELETE FROM people WHERE name = 'akmal';";
```

```
Statement statement = connection.createStatement();  
statement.executeUpdate(query);  
connection.commit();
```


Relational ...



... MySQL is actually a better NoSQL than most, if it's used as a NoSQL engine ...^[1]

... horizontally sharded MySQL data layer that allowed infinite horizontal scale.^[2]

... we decided to build our own simple, sharded datastore on top of MySQL.^[3]

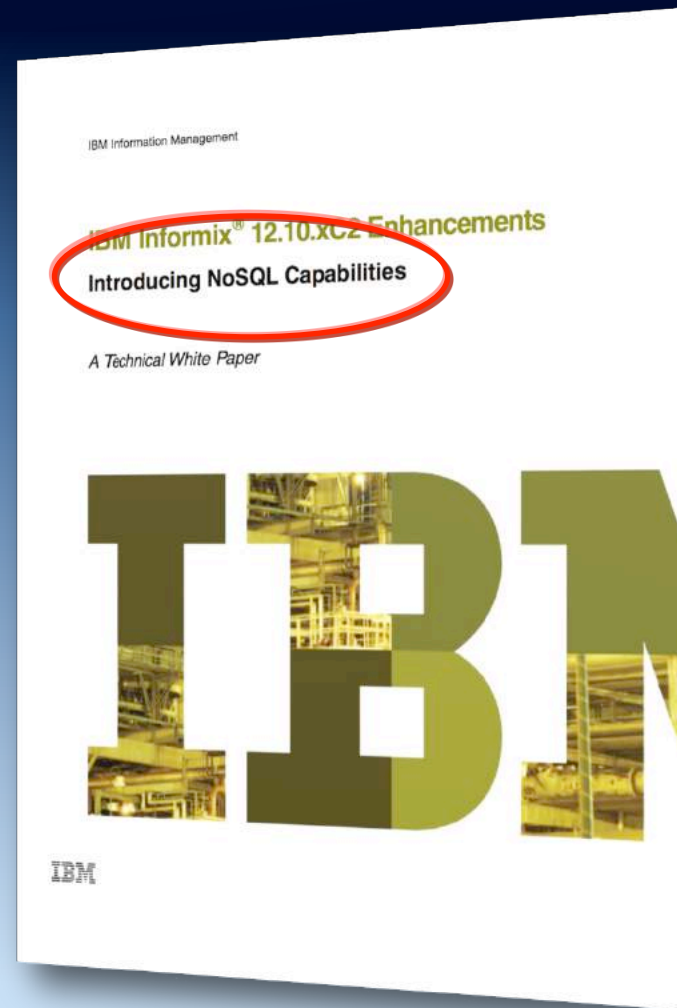
[1] <http://stackshare.io/wix/scaling-wix-to-60m-users---from-monolith-to-microservices/>

[2] <http://www.techrepublic.com/article/etsy-goes-retro-to-scale/>

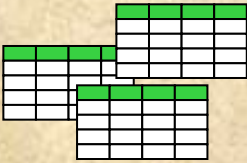
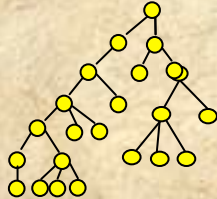
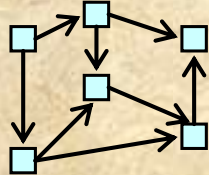
[3] <https://eng.uber.com/mezzanine-migration/>

Relational

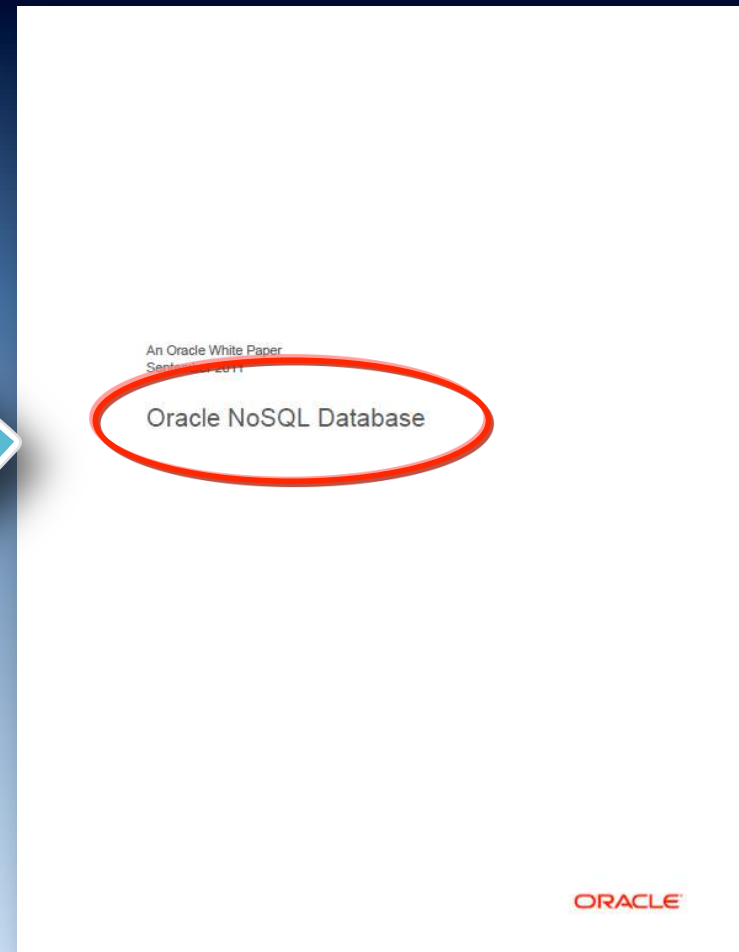
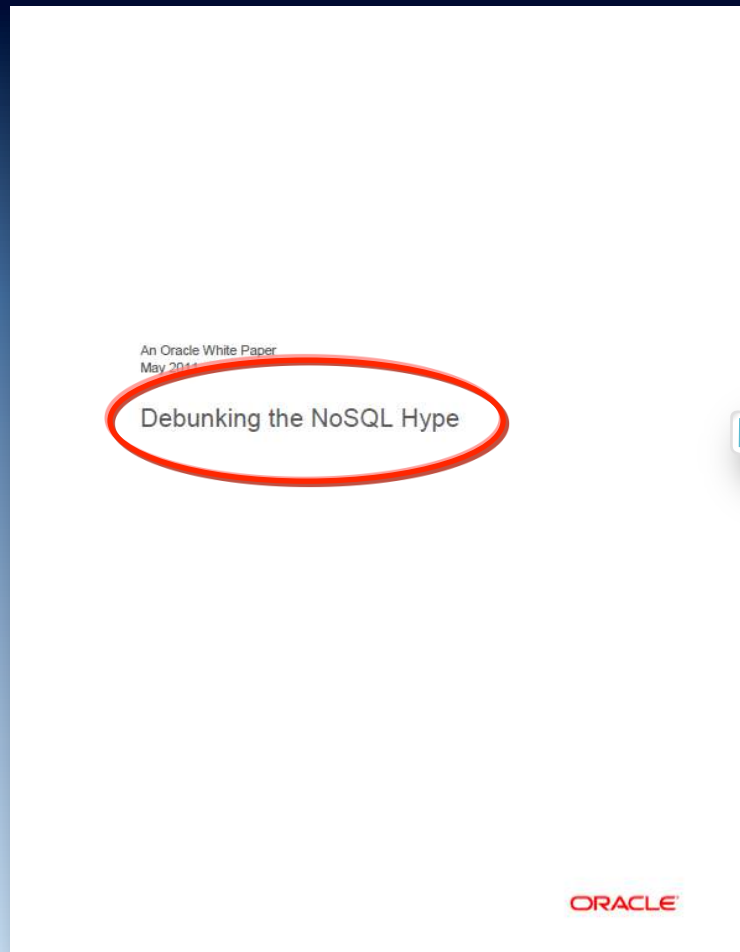
- Vendors adding NoSQL capabilities
 - Documents (JSON)
 - Linked data (RDF)



Relational vs. XML vs. RDF

Relational	XML	RDF
Tables 	Trees 	Graphs 
Flat, highly structured	Hierarchical data	Linked data
Rows in a table	Nodes in a tree	Triples describe links
Fixed schema	No or flexible schema	Highly flexible
SQL (ANSI/ISO)	XPath/XQuery (W3C)	SPARQL (W3C)

What about Oracle?



The meme changed (again)

Not
Only **SQL** → **No, SQL**





The rise of SQL ...



First they ignore you, then they laugh at you, then they fight you, then you win.

-- Mahatma Gandhi (disputed)

The rise of SQL

	Name	Example
	AQL	FOR ... IN ... FILTER ... RETURN
	CQL	SELECT ... FROM ... WHERE ...
	N1QL	SELECT ... FROM ... WHERE ...
		<code>db.collection.find({ ... })</code>

But ...



The bottom line here is to train your developers into understanding that even if it looks like SQL and quacks like SQL, if it's on a NoSQL database then it isn't SQL.

-- Andrew Cobley

And ...



... programmers have no idea what is going on behind the SQL façade, and, as a result, create programs that are wildly inefficient, far less efficient than the equivalent program in a traditional relational database.

-- Moshe Kranc

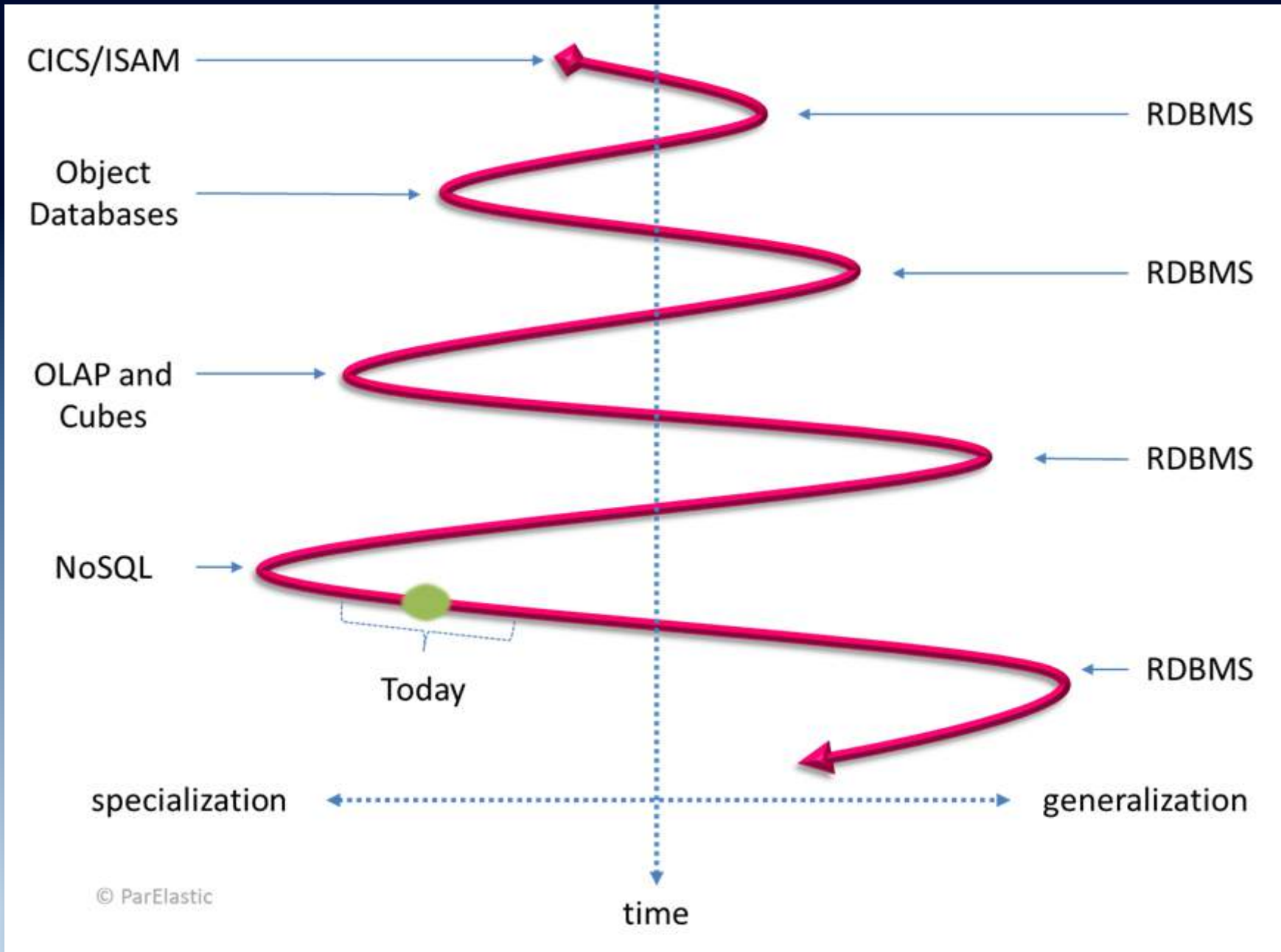
Summary



“The Time Tunnel”



Source: Shutterstock Image ID 135864122



Source: ParElastic, used with permission

History repeats



Those who cannot remember the past are condemned to repeat it.

-- George Santayana

Relational does NoSQL



Often the overhead of managing data in multiple databases is more than the advantages of the other store being faster.

You can do “NoSQL” inside and around a hackable database like PostgreSQL, not just as a separate one.

-- Hannu Krosing

“MySQL is web scale”

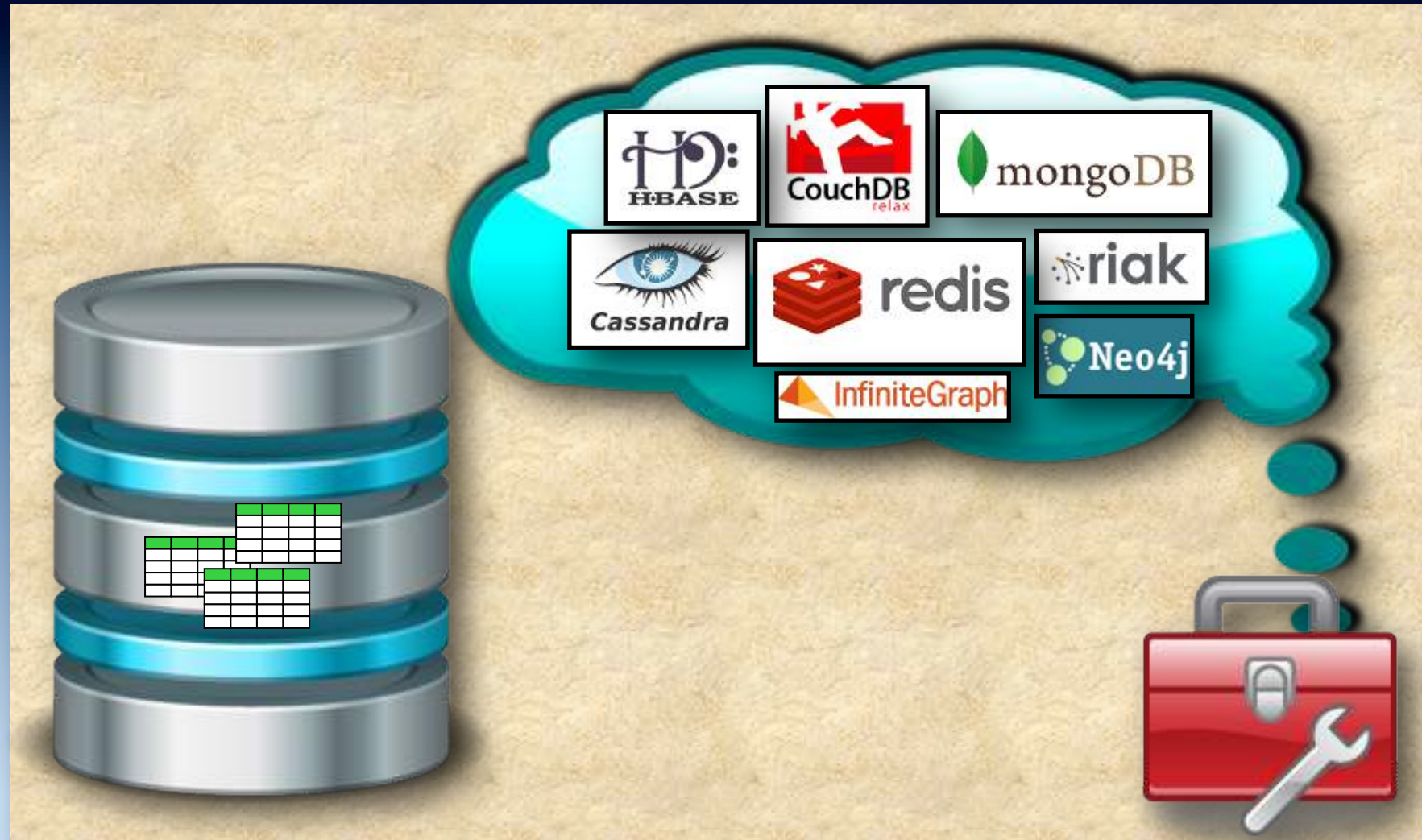


- Collaboration between Alibaba, Facebook, Google, LinkedIn and Twitter
- Adding more features to MySQL, specific to deployments in large-scale environments

Structured vs. unstructured



Relational vs. NoSQL toolbox



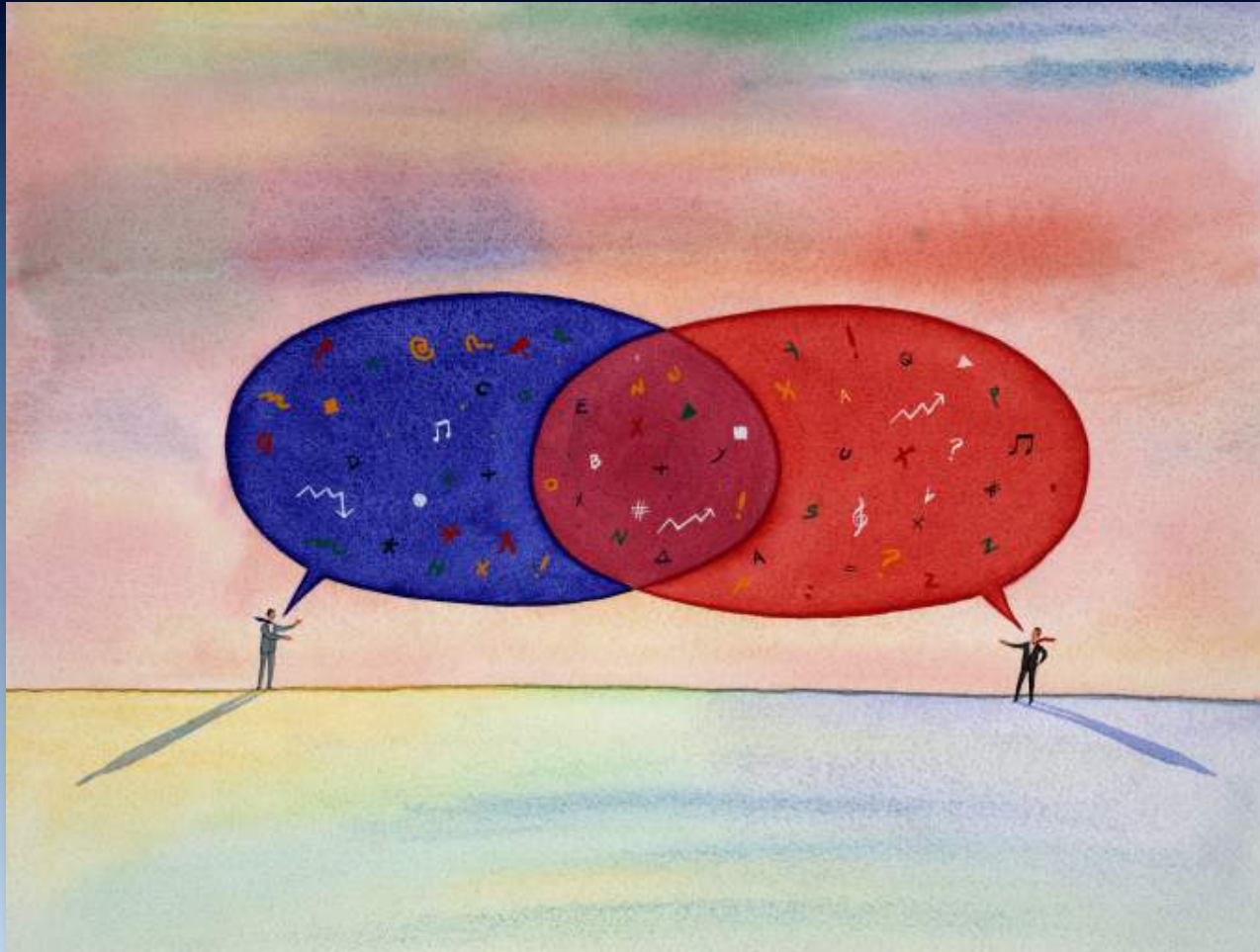
Relational vs. NoSQL ...



It is specious to compare NoSQL databases to relational databases; as you'll see, none of the so-called "NoSQL" databases have the same implementation, goals, features, advantages, and disadvantages. So comparing "NoSQL" to "relational" is really a shell game.

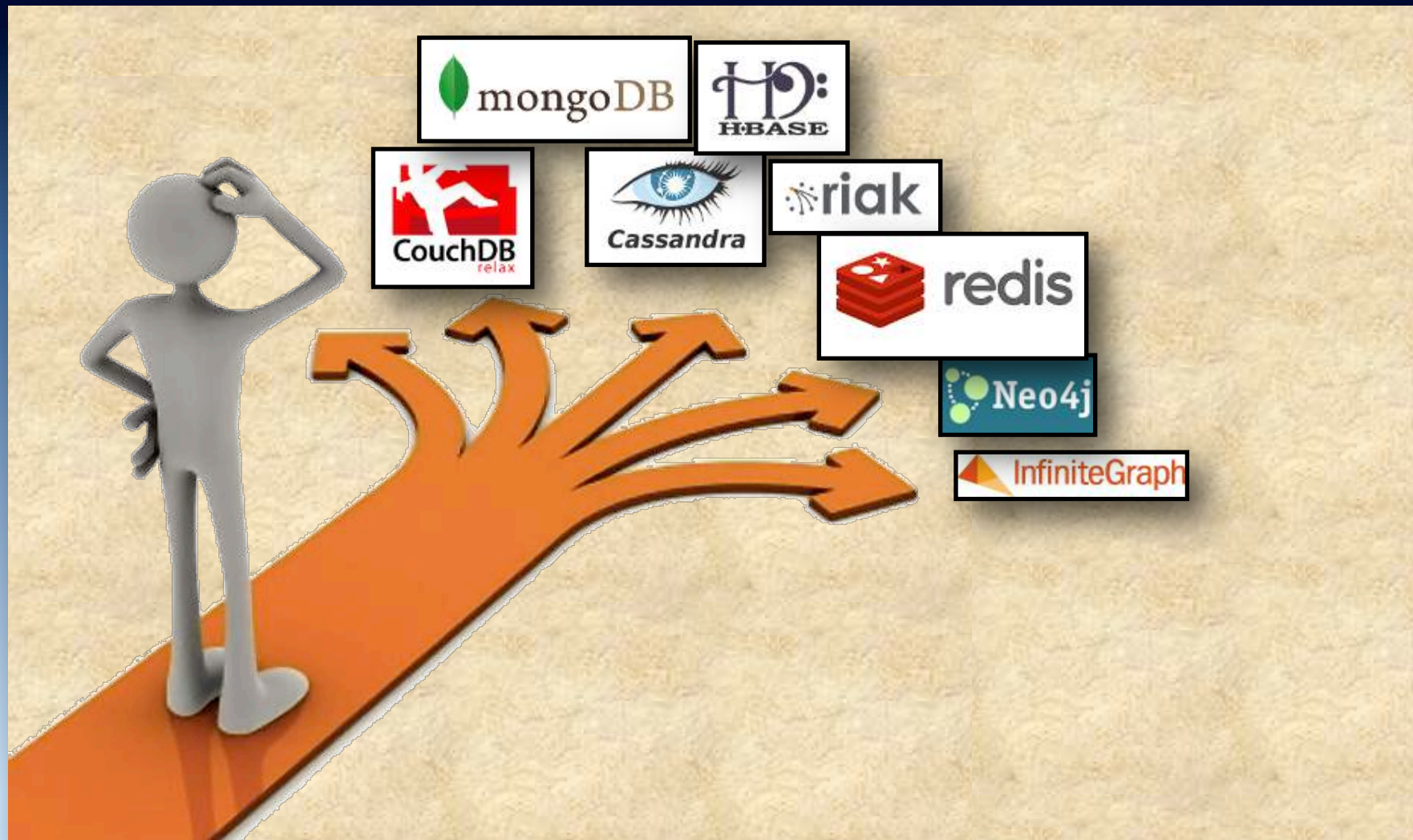
-- Eben Hewitt

Relational vs. NoSQL

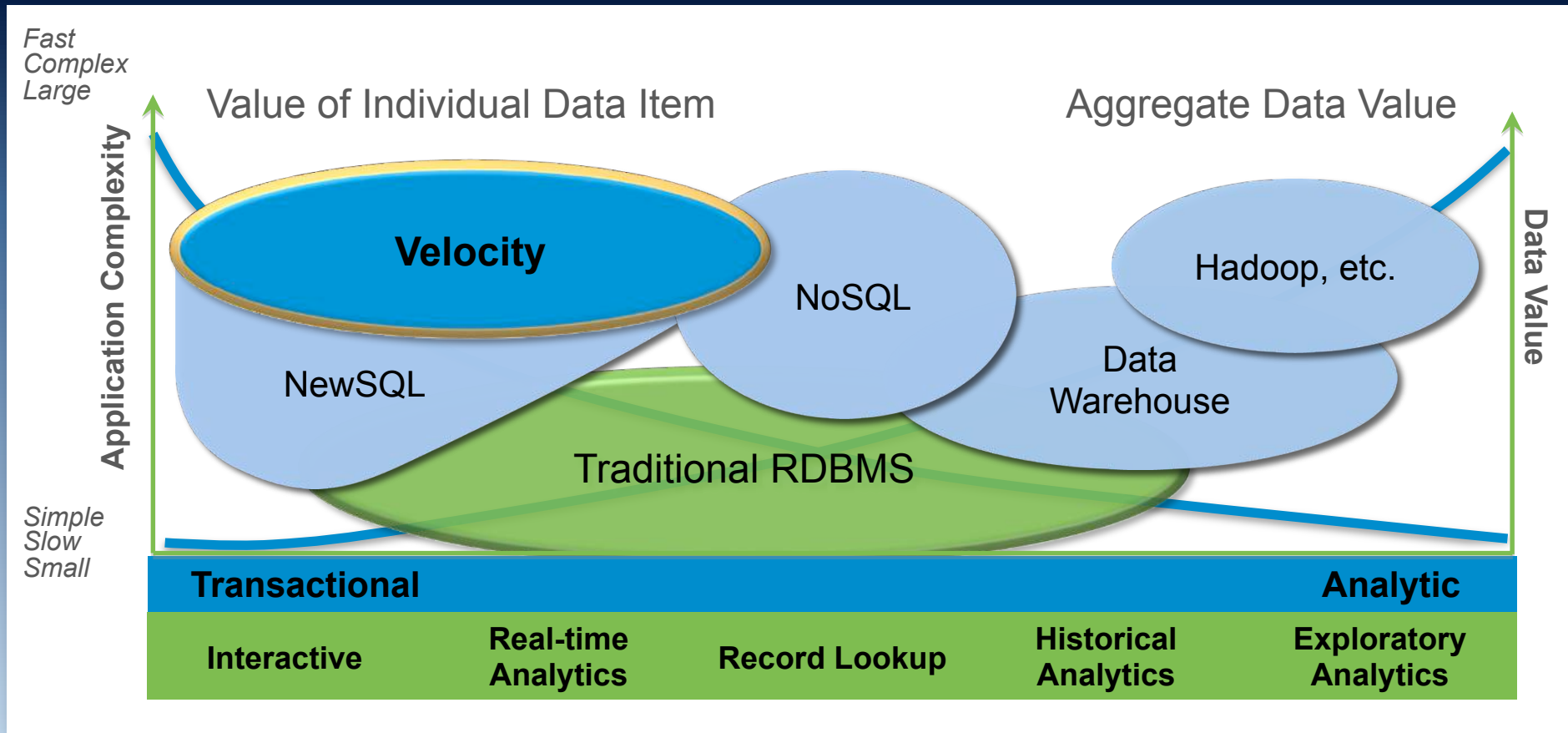


Source: Getty Image ID WCO_016

Choices, choices

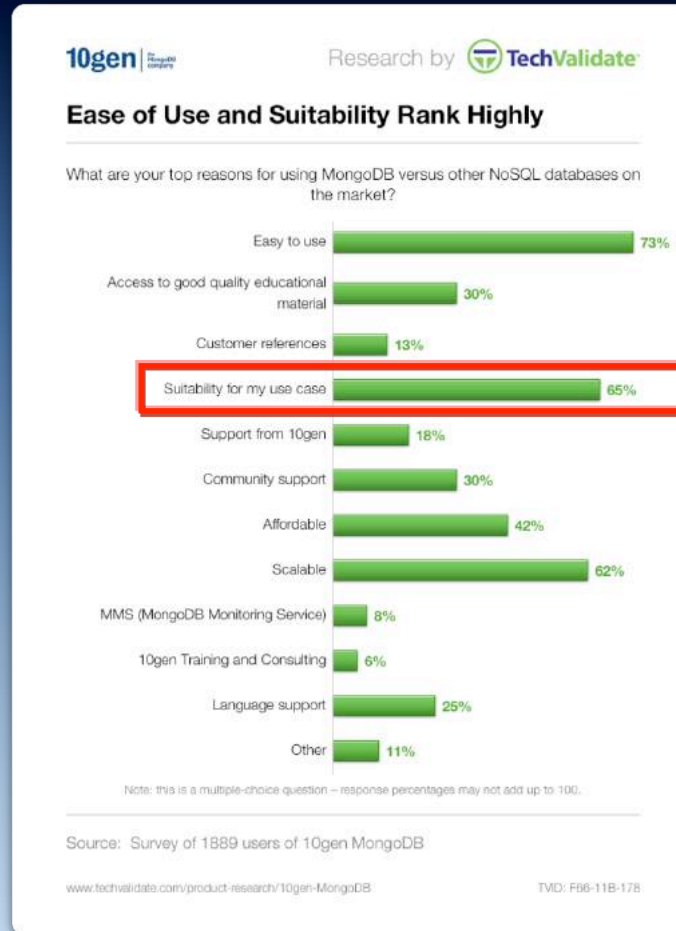


Navigating the DB universe



Source: VoltDB, used with permission

Understand your use case



Understand vendor-speak

What vendor says	What vendor means
The biggest in the world	The biggest one we've got
The biggest in the universe	The biggest one we've got
There is no limit to ...	It's untested, but we don't mind if you try it
A new and unique feature	Something the competition has had for ages
Currently available feature	We are about to start Beta testing
Planned feature	Something the competition has, that we wish we had too, that we might have one day
Highly distributed	International offices
Engineered for robustness	Comes in a tough box

Source: "Object Databases: An Evaluation and Comparison" Bloor Research (1994)

Vendor marketing example



*Really, really effective marketing masks
MongoDB's shortcomings...*

-- Robert Roland

Really effective marketing not unique to NoSQL



I would have made Oracle do serious quality control and not confuse future tense and present tense with regard to product features.

-- Mike Stonebraker

“Foundation”



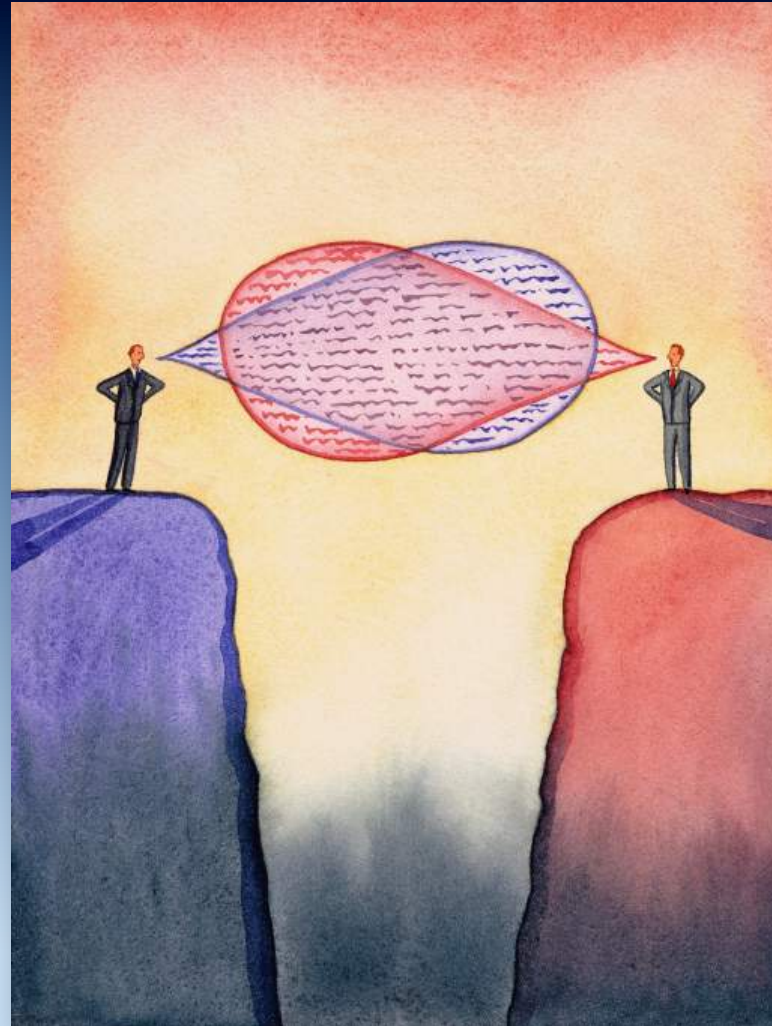
... there is a branch of human knowledge known as symbolic logic ... When Holk, after two days of steady work, succeeded in eliminating meaningless statements, vague gibberish, useless qualifications - in short, all the goo and dribble - he found he had nothing left. Everything canceled out.

-- Isaac Asimov

Understand the risks



The great debate ...



Source: Getty Image ID WCO_011

The great debate ...



About every ten years or so, there is a “great debate” between, on the one hand, those who see the problem of data modelling through a more or less relational lens, and on the other, a noisier set of “refuseniks” who have a hot new thing to promote. The debate usually goes like this:

The great debate ...



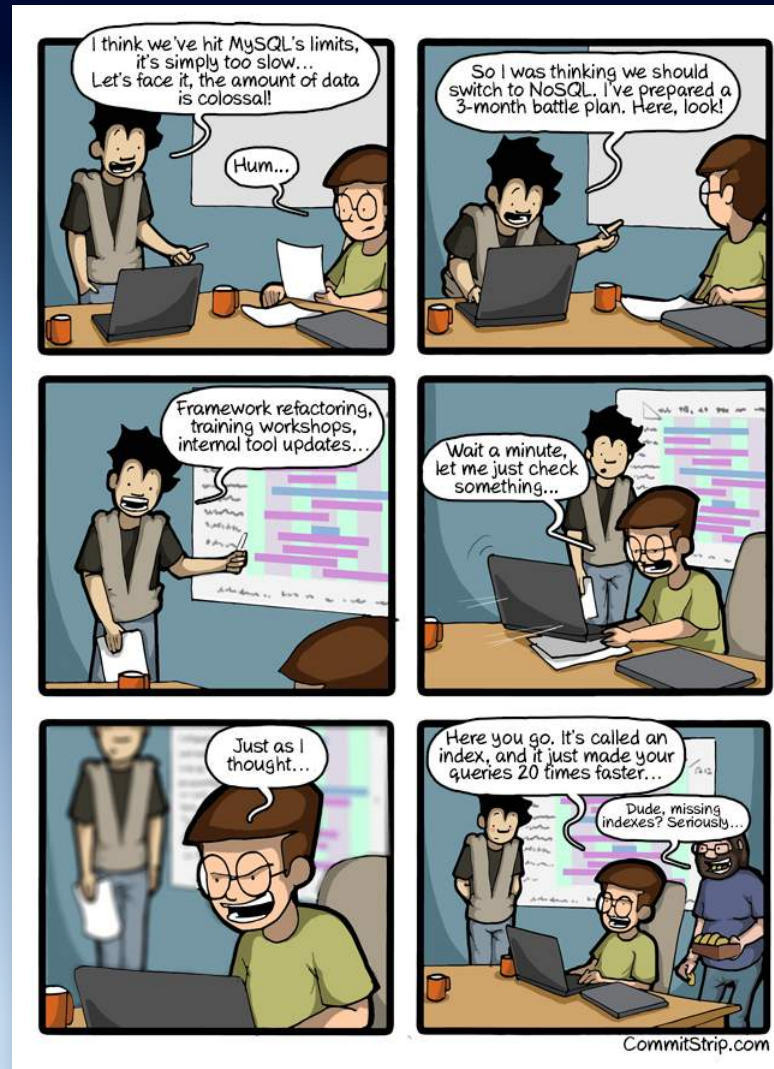
Refuseniks: Hah! You relational people with your flat tables and silly query languages! You are so unhip! You simply cannot deal with the problem of [INSERT NEW THING HERE]. With an [INSERT NEW THING HERE]-DBMS we will finish you, and grind your bones into dust!

The great debate



***R-people:** You make some good points. But unfortunately a) there is an enormous amount of money invested in building scalable, efficient and reliable database management products and no one is going to drop all of that on the floor and b) you are confusing DBMS engineering decisions with theoretical questions. We plan to incorporate the best of these ideas into our products.*

The problem is not the tool itself



Source: CommitStrip, used with permission

It's the people ...



... MongoDB Day London ... the problem is the people! They all talk like this:

- 1. Some problem that just doesn't really exist (or hasn't existed for a very long time) with relational databases*
 - 2. MongoDB*
 - 3. Profit!*
- Gaius Hammond*

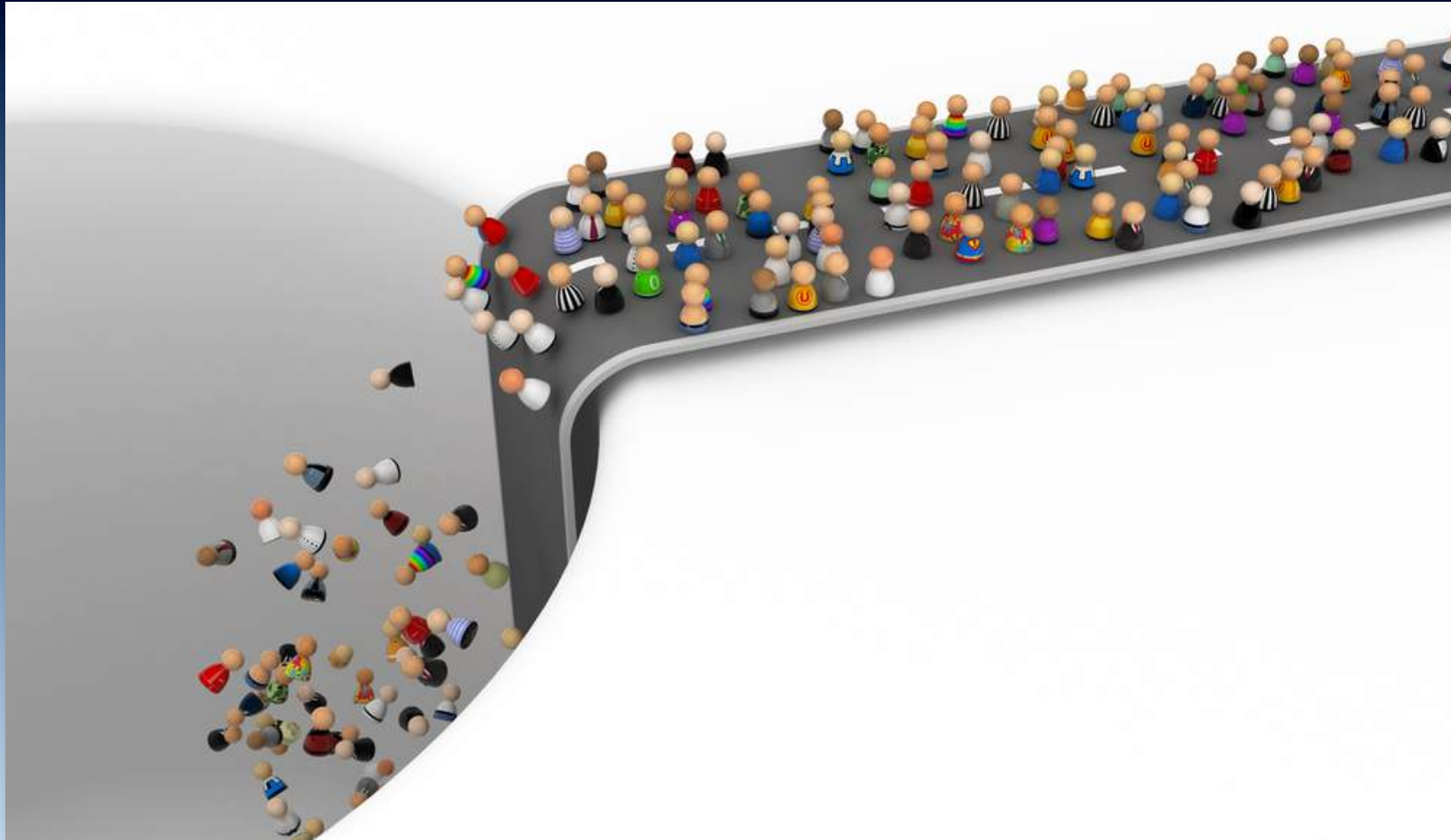
It's the people



... most of the business people driving the Big Data NoSQL databases are data management illiterate; don't recognize the lack of NoSQL data management facilities ... and don't know anything about availability, referential integrity and normalized data designs.

-- Dave Beulke

Don't be a Lemming



Source: Shutterstock Image ID 34566709

Limitations of NoSQL

- Lack of standardized or well-defined semantics
 - Transactions? Isolation levels?
- Reduced consistency for performance and scalability
 - “Eventual consistency”
 - “Soft commit”
- Limited forms of access, e.g. often no joins, etc.
- Proprietary interfaces
- Large clusters, failover, etc.?
- Security?

Hurdles to NoSQL adoption

- Immaturity of existing systems
- Lack of training and knowledge
- Too many choices
- Lack of mature tools
- The need for more use cases

Future directions

- Internal polyglot support (polymorphic?)
- Multi-model systems
- Google F1-inspired systems
 - “Can you have a scalable database without going NoSQL? Yes.”
- Further support for NoSQL in Relational
- DBaaS



Final thoughts



We are clearly in the phase of a new technology adoption in which the category is hyped, its benefits over-promised, its limitations poorly understood, and its value oversold.

-- Tim Berglund

There will be harmony



Source: Shutterstock Image ID 73418620

DAILY
200000

CITY EXAMINER

LATE EDITION

BREAKING: ANIMAL CHAUDHRI GIVES ANOTHER AWESOME NOSQL PRESENTATION!

NEWS UPDATE

CHAS DID ANOTHER presentation in
times. Chaudhri did a
great job. He showed us
how to use the database
to store data. It was
very interesting.

SPECIAL REPORT



Island Vacations

One week. That's all it takes to
escape the city and enjoy
the sun and sand. It's
perfect for anyone who
wants to relax and
recharge.








GOSSIP

Dear William, what is it
about you? You are
so interesting. I
love to hear about
your life. It's so
different from
mine.

Contact details



Find me on

-  <http://www.linkedin.com/in/akmalchaudhri/>
-  <http://twitter.com/akmalchaudhri/>
-  <http://www.quora.com/Akmal-Chaudhri/>
-  <http://www.facebook.com/akmal.chaudhri/>
-  <http://plus.google.com/+AkmalChaudhri/>
-  <http://www.slideshare.net/VeryFatBoy/>
-  <http://www.youtube.com/VeryFatBoyVideos/>



Akmal B. Chaudhri
firstname.lastname@live.com

Questions?



Source: Shutterstock Image ID 194875901

```
{"thank": "You"}
```

Resources



Recommended reading ...

- Choosing the right NoSQL database for the job: a quality attribute evaluation
 - <http://www.journalofbigdata.com/content/2/1/18/>
- Gartner Magic Quadrant for Operational Database Management Systems (2015)
 - <https://info.microsoft.com/CO-SQL-CNTNT-FY16-09Sep-14-MQOperational-Register.html>

Recommended reading

- Learn to stop using shiny new things and love MySQL
 - <https://engineering.pinterest.com/blog/learn-stop-using-shiny-new-things-and-love-mysql/>
- MongoDB Days
 - <https://gaiustech.wordpress.com/2013/04/13/mongodb-days/>

History ...

- First NoSQL meetup
 - <http://nosql.eventbrite.com/>
 - <http://blog.oskarsson.nu/post/22996139456/nosql-meetup>
- First NoSQL meetup debrief
 - <http://blog.oskarsson.nu/post/22996140866/nosql-debrief>
- First NoSQL meetup photographs
 - <http://www.flickr.com/photos/russss/sets/72157619711038897/>

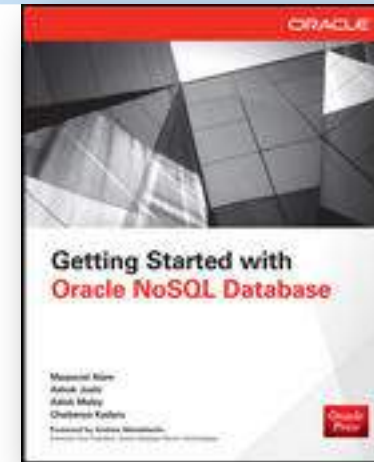
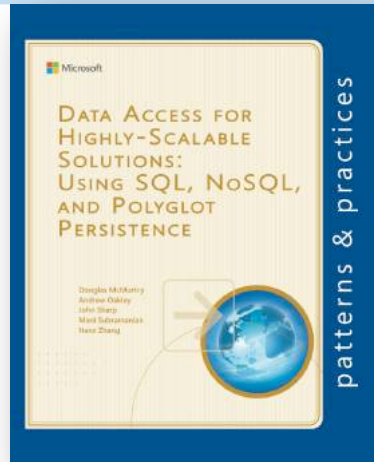
History

- Codd's Relational Vision - Has NoSQL Come Full Circle?
 - <http://www.opensourceconnections.com/2013/12/11/codds-relational-vision-has-nosql-come-full-circle/>

Web sites

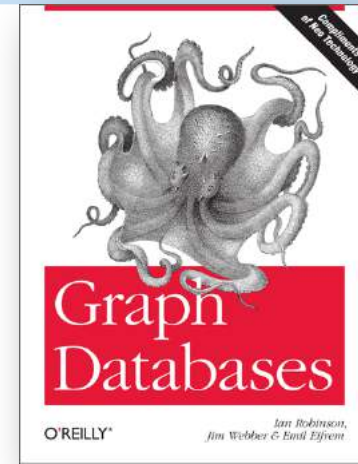
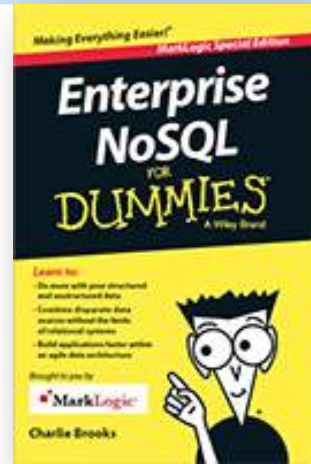
- NoSQL Databases and Polyglot Persistence: A Curated Guide
 - <http://nosql.mypopescu.com/>
- NoSQL: Your Ultimate Guide to the Non-Relational Universe!
 - <http://nosql-database.org/>

Free books ...



- Data Access for Highly-Scalable Solutions: Using SQL, NoSQL, and Polyglot Persistence
 - <http://www.microsoft.com/en-us/download/details.aspx?id=40327>
- Getting Started with Oracle NoSQL Database
 - <http://books.mcgraw-hill.com/ebookdownloads/NoSQL/>

Free books ...



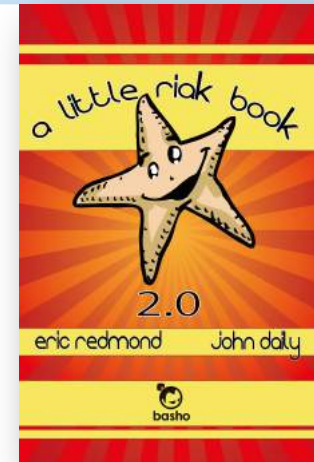
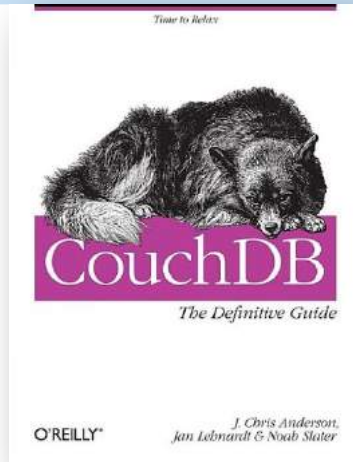
- Enterprise NoSQL for Dummies
 - <http://www.nosqlfordummies.com/>
- Graph Databases
 - <http://www.graphdatabases.com/>

Free books ...



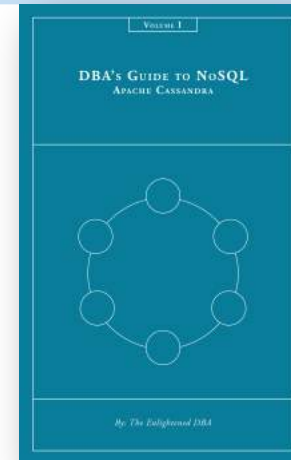
- The Little MongoDB Book
 - <http://openmymind.net/mongodb.pdf>
- The Little Redis Book
 - <http://openmymind.net/redis.pdf>

Free books ...



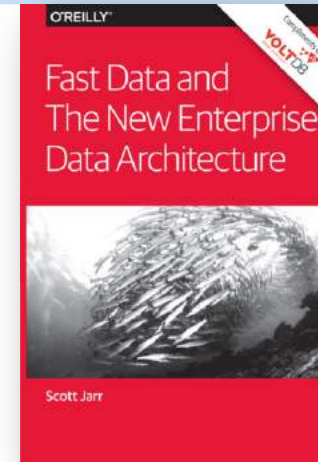
- CouchDB: The Definitive Guide
 - <http://guide.couchdb.org/>
- A Little Riak Book
 - https://github.com/coderoshi/little_riak_book/

Free books ...



- Understanding The Top 5 Redis Performance Metrics
 - <https://www.datadoghq.com/wp-content/uploads/2013/09/Understanding-the-Top-5-Redis-Performance-Metrics.pdf>
- DBA's Guide to NoSQL
 - <https://www.smashwords.com/books/view/479798/>

Free books



- Mastering Hazelcast
 - <http://hazelcast.com/resources/mastering-hazelcast/>
- Fast Data and the New Enterprise Data Architecture
 - <http://voltdb.com/fast-data-and-new-enterprise-data-architecture/>

Free training ...



- MongoDB
 - <https://university.mongodb.com/>

Free training ...

- Aerospike
 - [http://www.aerospike.com/training/<administration | development>/online/](http://www.aerospike.com/training/<administration|development>/online/)
- Cassandra
 - <https://academy.datastax.com/>
- Couchbase
 - <https://training.couchbase.com/online>

Free training

- Neo4j
 - http://www.neo4j.org/learn/online_course/
- OrientDB
 - <http://www.orienttechnologies.com/getting-started/>

Articles ...

- The State of NoSQL
 - <http://www.infoq.com/articles/State-of-NoSQL/>
- An Introduction to NoSQL Patterns
 - <http://architects.dzone.com/articles/introduction-nosql-patterns>
- The NoSQL Advice I Wish Someone Had Given Me
 - <http://sql.dzone.com/articles/nosql-advice-i-wish-someone>

Articles ...

- Why is the NoSQL choice so difficult?
 - <http://www.itworld.com/article/2696615/big-data/why-is-the-nosql-choice-so-difficult-.html>
- NoSQL is a no go once again
 - <http://www.itworld.com/article/2696893/big-data/nosql-is-a-no-go-once-again.html>

Articles

- Why horizontal scalability shouldn't be a focus for software startups
 - <http://www.itworld.com/article/2984271/development/why-horizontal-scalability-shouldnt-be-a-focus-for-software-startups.html>

Free reports ...

- A deep dive into NoSQL: A complete list of NoSQL databases
 - <http://www.bigdata-madesimple.com/a-deep-dive-into-nosql-a-complete-list-of-nosql-databases/>
- Deconstructing NoSQL
 - <http://whitepapers.dataversity.net/content37165/>
- Dzone's Guide to Database & Persistence Management
 - <https://dzone.com/guides/database-persistence-management>

Free reports ...

- Gartner Magic Quadrant for Operational Database Management Systems (2013)
 - <http://oracledbacr.blogspot.co.uk/2014/01/magic-quadrant-for-operational-database.html>
- Gartner Magic Quadrant for Operational Database Management Systems (2015)
 - <https://info.microsoft.com/CO-SQL-CNTNT-FY16-09Sep-14-MQOperational-Register.html>

Free reports ...

- Five Data Persistence Dilemmas That Will Keep CIOs Up at Night
 - <http://www1.memsql.com/gartner-cio-report/>
- Critical Capabilities for Operational Database Management Systems
 - <http://go.nuodb.com/gartner-critical-capabilities.html>
- When to Use New RDBMS Offerings in a Dynamic Data Environment
 - <http://go.nuodb.com/avant-garde-databases.html>

Free reports ...

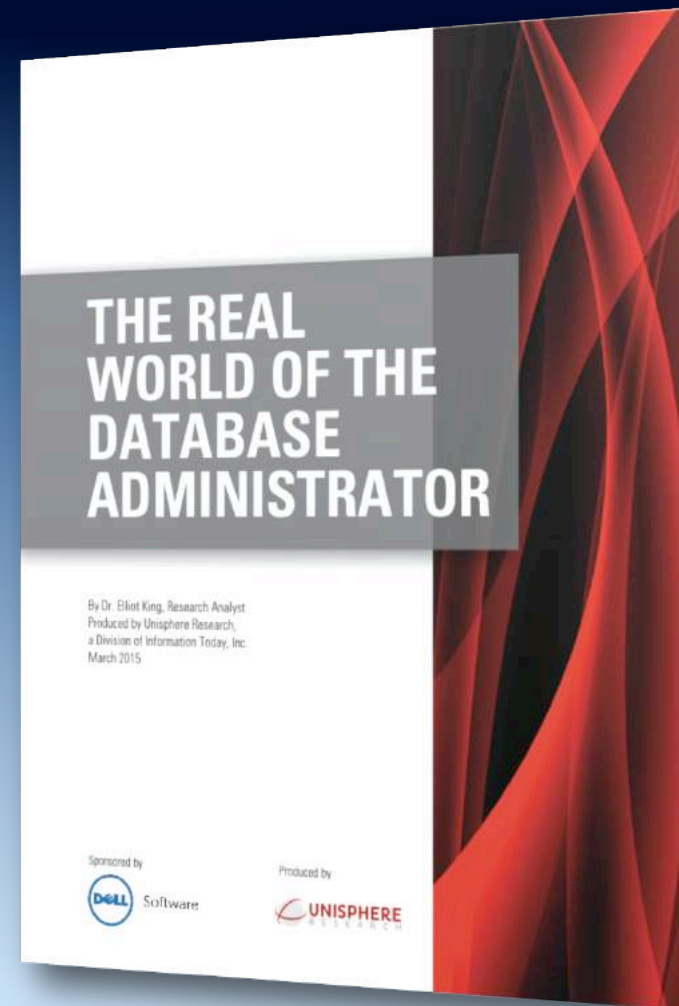
- The Forrester Wave™: NoSQL Key-Value Databases, Q3 2014
 - <https://www.mapr.com/forrester-wave-hadoop-nosql-key-value-databases>
- The Forrester Wave™: NoSQL Document Databases, Q3 2014
 - <http://info.marklogic.com/forrester-wave.html>
- Forrester Ranks the NoSQL Database Vendors
 - <http://www.datanami.com/2014/10/03/forrester-ranks-nosql-database-vendors/>

Free reports ...

- The Forrester Wave™: In-Memory Database Platforms, Q3 2015
 - <http://www1.memsql.com/forrester/>

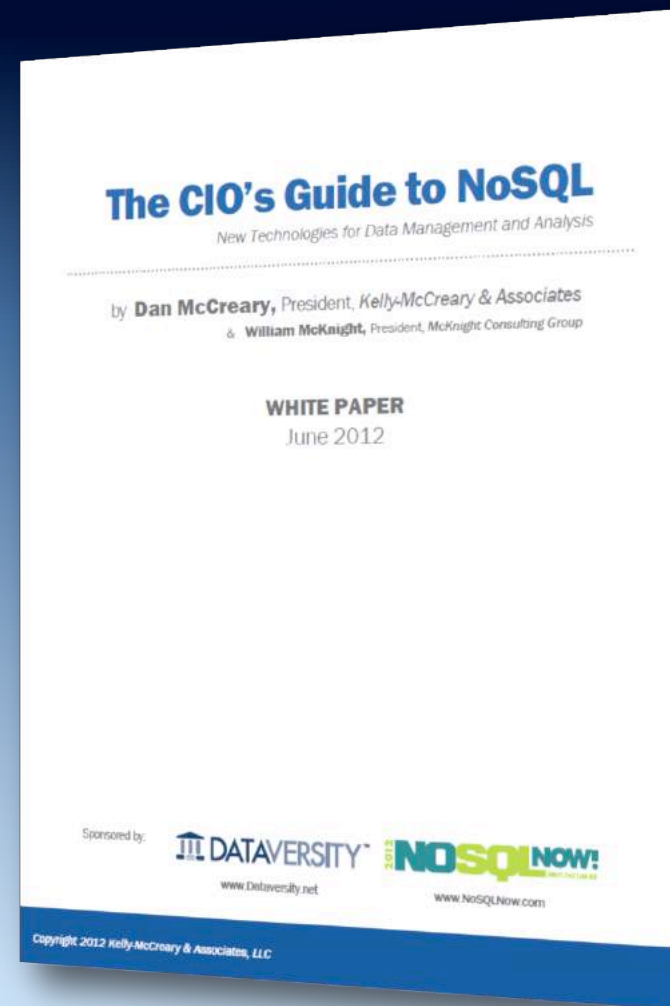
Free reports

- The Real World of The Database Administrator
 - <https://software.dell.com/whitepaper/the-real-world-of-the-database-administrator-875469/>



White papers

- The CIO's Guide to NoSQL
 - <http://documents.dataversity.net/whitepapers/the-cios-guide-to-nosql.html>



Vendor funding ...

- Visualizing the \$1bn+ VC investment in Hadoop and NoSQL
 - http://blogs.the451group.com/information_management/2013/12/17/visualizing-the-1bn-vc-investment-in-hadoop-and-nosql/
- Hadoop vs. NoSQL - Which Big Data Technology Has Raised More Funding?
 - <http://www.cbinsights.com/blog/hadoop-nosql-venture-capital-funding/>

Vendor funding

- The NoSQLNow conference in San Jose this week
 - <http://swtrends.wordpress.com/2014/08/22/the-nosqlnow-conference-in-san-jose-this-week/>
- NoSQL market frames larger debate: Can open source be profitable?
 - <http://siliconangle.com/blog/2015/03/19/nosql-market-frames-larger-debate-can-open-source-be-profitable/>

Brewer's CAP "Theorem" ...

- Towards Robust Distributed Systems
 - <http://www.cs.berkeley.edu/~brewer/cs262b-2004/PODC-keynote.pdf>
- Deconstructing the 'CAP theorem' for CM and DevOps
 - http://markburgess.org/blog_cap.html
- NoCAP Or, Achieving Scalability Without Compromising on Consistency
 - <http://www.gigaspace.com/system/files/private/resource/NoCAPfinal0711.pdf>

Brewer's CAP "Theorem" ...

- Brewer's CAP Theorem
 - <http://www.julianbrowne.com/article/viewer/brewers-cap-theorem>
- Confused CAP Arguments
 - <http://www.stucharlton.com/blog/archives/2010/10/confused-cap-arguments.html>
- Please stop calling databases CP or AP
 - <https://martin.kleppmann.com/2015/05/11/please-stop-calling-databases-cp-or-ap.html>

Brewer's CAP "Theorem"

- The CAP theorem series
 - <http://blog.thislongrun.com/2015/03/the-cap-theorem-series.html>

Data consistency

- Replicated Data Consistency Explained Through Baseball
 - <http://research.microsoft.com/apps/pubs/default.aspx?id=206913>
- Distributed Algorithms in NoSQL Databases
 - <https://highlyscalable.wordpress.com/2012/09/18/distributed-algorithms-in-nosql-databases/>

Product selection ...

- 101 Questions to Ask When Considering a NoSQL Database
 - <http://highscalability.com/blog/2011/6/15/101-questions-to-ask-when-considering-a-nosql-database.html>
- 35+ Use Cases for Choosing Your Next NoSQL Database
 - <http://highscalability.com/blog/2011/6/20/35-use-cases-for-choosing-your-next-nosql-database.html>

Product selection ...

- NoSQL Data Modeling Techniques
 - <http://highlyscalable.wordpress.com/2012/03/01/nosql-data-modeling-techniques/>
- Choosing a NoSQL data store according to your data set
 - <http://00f.net/2010/05/15/choosing-a-nosql-data-store-according-to-your-data-set/>
- The Right Database for Your Use Case
 - <http://mpron.github.io/the-right-database-for-your-use-case/>

Product selection ...

- NoSQL Options Compared: Different Horses for Different Courses
 - <http://www.slideshare.net/tazija/nosql-options-compared/>
- The NoSQL Technical Comparison Report: Cassandra (DataStax), MongoDB, and Couchbase Server
 - <http://www.altoros.com/nosql-tech-comparison-cassandra-mongodb-couchbase.html>

Product selection ...

- The Solutions Architect's Guide to Choosing a (NoSQL) Data Store
 - <http://bogdanbocse.com/2014/12/the-solutions-architects-guide-to-choosing-a-nosql-data-store-process-overview/>
 - <http://bogdanbocse.com/2014/12/the-solutions-architects-guide-to-choosing-a-nosql-data-store-analyze-the-requirements-of-your-ideal-solutions/>

Product selection

- Design Assistant for NoSQL Technology Selection
 - <http://dl.acm.org/citation.cfm?id=2751494>

Short product overviews

- Cassandra vs MongoDB vs CouchDB vs Redis vs Riak vs HBase vs Couchbase vs Neo4j vs Hypertable vs Elasticsearch vs Accumulo vs VoltDB vs Scalaris comparison
 - <http://kkovacs.eu/cassandra-vs-mongodb-vs-couchdb-vs-redis/>
- vsChart.com
 - <http://vschart.com/list/database/>

Case studies ...

- Choosing a NoSQL: A Real-Life Case
 - <http://www.slideshare.net/VolhaBanadyseva/10-ss-choosing-a-nosql-database/>
- From 1000/day to 1000/sec: The Evolution of Incapsula's BIG DATA System
 - <http://www.slideshare.net/Incapsula/surge2014/>
- Providence: Failure Is Always an Option
 - <http://jasonpunyon.com/blog/2015/02/12/providence-failure-is-always-an-option/>

Case studies

- NoSQL Data Store Technologies
 - <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA611676>

NoSQL alternatives ...

- Learn to stop using shiny new things and love MySQL
 - <https://engineering.pinterest.com/blog/learn-stop-using-shiny-new-things-and-love-mysql/>
- Etsy goes retro to scale big data
 - <http://www.techrepublic.com/article/etsy-goes-retro-to-scale/>
- Project Mezzanine: The Great Migration
 - <https://eng.uber.com/mezzanine-migration/>

NoSQL alternatives ...

- Our Race for a New Database
 - <https://eng.uber.com/schemaless-part-one/>
- Schemaless Synopsis
 - <https://eng.uber.com/schemaless-part-two/>
- Using Triggers On Schemaless, Uber Engineering's Datastore Using MySQL
 - <https://eng.uber.com/schemaless-part-three/>

NoSQL alternatives

- Best practices for scaling with DevOps and microservices
 - <http://techbeacon.com/how-wix-scaled-devops-microservices>
- Scaling Wix to 60M Users - From Monolith to Microservices
 - <http://stackshare.io/wix/scaling-wix-to-60m-users---from-monolith-to-microservices/>
- MySQL is a Great NoSQL Database
 - <https://dzone.com/articles/mysql-is-a-great-nosql-1>

High-profile MySQL web sites

- Facebook
 - <http://www.mysql.com/customers/view/?id=757>
- Twitter
 - <http://www.mysql.com/customers/view/?id=951>
- Tumblr
 - <http://www.mysql.com/customers/view/?id=1186>
- Wikipedia
 - <http://www.mysql.com/customers/view/?id=663>

Negative NoSQL comments ...

- MongoDB is to NoSQL like MySQL to SQL - in the most harmful way
 - <http://use-the-index-luke.com/blog/2013-10/mysql-is-to-sql-like-mongodb-to-nosql>
- The Genius and Folly of MongoDB
 - <http://nyeggen.com/post/2013-10-18-the-genius-and-folly-of-mongodb/>
- Why You Should Never Use MongoDB
 - <http://www.sarahmei.com/blog/2013/11/11/why-you-should-never-use-mongodb/>

Negative NoSQL comments ...

- Failing with MongoDB
 - <http://blog.schmichael.com/2011/11/05/failing-with-mongodb/>
 - <https://speakerdeck.com/robotadam/postgres-at-urban-airship/>
- A Year with MongoDB
 - <http://blog.kiip.me/engineering/a-year-with-mongodb/>
 - <https://speakerdeck.com/mitsuhiko/a-year-of-mongodb/>

Negative NoSQL comments ...

- Why MongoDB Never Worked Out at Etsy
 - <http://mcfunley.com/why-mongodb-never-worked-out-at-etsy/>
- A post you wish to read before considering using MongoDB for your next app
 - <http://longtermlaziness.wordpress.com/2012/08/24/a-post-you-wish-to-read-before-considering-using-mongodb-for-your-next-app/>

Negative NoSQL comments ...

- Goodbye, CouchDB
 - <http://sauceio.com/index.php/2012/05/goodbye-couchdb/>
- Don't use NoSQL
 - <https://speakerdeck.com/roidrage/dont-use-nosql/>
 - <http://vimeo.com/49713827/>
- The SQL and NoSQL Effects: Will They Ever Learn?
 - <http://www.dbdebunk.com/2015/07/the-sql-and-nosql-effects-will-they.html>

Negative NoSQL comments ...

- Do Developers Use NoSQL Because They're Too Lazy to Use RDBMS Correctly?
 - <http://architects.dzone.com/articles/do-developers-use-nosql>
 - <http://gaiustech.wordpress.com/2013/04/13/mongodb-days/>
- The parallels between NoSQL and self-inflicted torture
 - <http://www.parelastic.com/blog/parallels-between-nosql-and-self-inflicted-torture/>

Negative NoSQL comments

- 7 hard truths about the NoSQL revolution
 - <http://www.infoworld.com/article/2617405/nosql/7-hard-truths-about-the-nosql-revolution.html>
- Google goes back to the future with SQL F1 database
 - http://www.theregister.co.uk/2013/08/30/google_f1_deepdive/
- What's left of NoSQL?
 - <http://use-the-index-luke.com/blog/2013-04/whats-left-of-nosql>

Gotchas ...

- Five Ways Open Source Databases Are Limited
 - <http://www.datanami.com/2015/09/03/five-ways-open-source-databases-are-limited/>
- Operations costs are the Achilles' heel of NoSQL
 - <http://www.computerworld.com/article/2997183/cloud-storage/operations-costs-are-the-achilles-heel-of-nosql.html>

Gotchas ...

- Broken by Design: MongoDB Fault Tolerance
 - <http://hackingdistributed.com/2013/01/29/mongo-ft/>
- Things they don't tell you about MongoDB
 - <http://www.itexto.com.br/devkico/en/?p=44>
- MongoDB Gotchas & How To Avoid Them
 - <http://rsmith.co/2012/11/05/mongodb-gotchas-and-how-to-avoid-them/>

Gotchas

- Top 5 syntactic weirdnesses to be aware of in MongoDB
 - <http://devblog.me/wtf-mongo>
- This Team Used Apache Cassandra... You Won't Believe What Happened Next
 - <http://blog.parseely.com/post/1928/cass/>

NoSQL to Relational ...

- MongoDB to MySQL (Aadhar)
 - <http://techcrunch.com/2013/12/06/inside-indias-aadhar-the-worlds-biggest-biometrics-database/>
- MongoDB to MySQL (Diaspora)
 - <http://www.slideshare.net/sarahmei/taking-diaspora-from-mongodb-to-mysql-rubyconf-2011/>
- Redis to MySQL (OpenSource Connections)
 - <http://www.slideshare.net/AllThingsOpen/stop-worrying-love-the-sql-a-case-study/>

NoSQL to Relational ...

- MongoDB to PostgreSQL (Urban Airship)
 - <http://blog.schmichael.com/2011/11/05/failing-with-mongodb/>
- MongoDB to Postgres
 - <http://blog.testdouble.com/posts/2014-06-23-mongo-to-postgres.html>
- MongoDB to PostgreSQL (Errbit fork)
 - <https://github.com/errbit/errbit/issues/614/>

NoSQL to Relational ...

- MongoDB to PostgreSQL (Olery)
 - <http://developer.olery.com/blog/goodbye-mongodb-hello-postgresql/>
- NoSQL to PostgreSQL (Revolv)
 - <http://technosophos.com/2014/04/11/nosql-no-more.html>
- MongoDB to NuoDB (DropShip Commerce)
 - <http://searchdatamanagement.techtarget.com/feature/NewSQL-database-sends-NoSQL-technology-packing-at-logistics-exchange>

NoSQL to Relational

- RavenDB to SQL Server (Octopus)
 - <https://octopusdeploy.com/blog/3.0-switching-to-sql/>
- MongoDB to Vertica (Twin Prime)
 - <http://engineering.twinprime.com/sql-or-nosql/>

NoSQL to NoSQL ...

- MongoDB. This is not the database you are looking for.
 - <http://patrickmcfadin.com/2014/02/11/mongodb-this-is-not-the-database-you-are-looking-for/>
- MongoDB to Couchbase (Viber)
 - <http://www.slideshare.net/Couchbase/couchbaseatv2014couchbaseatviber/>
- MongoDB to HBase (Simply Measured)
 - <http://www.slideshare.net/RobertRoland2/rebuilding-22995359/>

NoSQL to NoSQL ...

- MongoDB to Cassandra (MetaBroadcast)
 - <http://www.slideshare.net/fredvdd/mongodb-to-cassandra/>
- MongoDB to Cassandra (SHIFT)
 - <http://www.slideshare.net/DataStax/shift-real-world-migration-from-mongo-db-to-cassandra-25970769/>
- MongoDB to Cassandra (FullContact)
 - <http://www.fullcontact.com/blog/mongo-to-cassandra-migration/>

NoSQL to NoSQL ...

- MongoDB to Cassandra (Shodan)
 - <http://planetcassandra.org/blog/post/mongodb-to-cassandra-a-developers-story/>
- MongoDB to Cassandra (Retailigence)
 - <http://planetcassandra.org/blog/post/retailigence-turns-to-apache-cassandra-after-returning-mysql-and-mongodb-for-scalable-location-based-shopping-api/>
- MongoDB to Neo4j (Shindig)
 - <http://seenickcode.com/switching-from-mongodb-to-neo4j/>

NoSQL to NoSQL ...

- MongoDB to Cloudant (Postmark)
 - <http://blog.postmarkapp.com/post/37338222496/bye-mongodb-hello-cloudant/>
- MongoDB to Cloudant (IBM)
 - <http://blog.ibmjstart.net/2015/08/05/porting-from-mongodb-to-cloudant-differences-in-design/>
- MongoDB to DynamoDB (Gummicube)
 - <https://www.codementor.io/devops/tutorial/handling-date-and-datetime-in-dynamodb/>

NoSQL to NoSQL

- Cassandra to DynamoDB (Tellybug)
 - <http://attentionshard.wordpress.com/2013/09/30/why-tellybug-moved-from-cassandra-to-amazon-dynamodb/>
- Redis to Cassandra (Instagram)
 - <http://planetcassandra.org/blog/post/cassandra-summit-2013-instagrams-shift-to-cassandra-from-redis-by-rick-branson/>

Security ...

- Abusing NoSQL Databases
 - <https://www.defcon.org/images/defcon-21/dc-21-presentations/Chow/DEFCON-21-Chow-Abusing-NoSQL-Databases.pdf>
- NoSQL, no security?
 - <http://www.slideshare.net/wurbanski/nosql-no-security/>
- NoSQL, No Injection!?
 - http://www.slideshare.net/wayne_armorize/nosql-no-sql-injections-4880169/

Security ...

- NoSQL, But Even Less Security
 - <http://blogs.adobe.com/asset/files/2011/04/NoSQL-But-Even-Less-Security.pdf>
- NoSQL Database Security
 - <http://pastconferences.auscert.org.au/conf2011/presentations/Louis%20Nyffenegger%20V1.pdf>
- Does NoSQL Mean No Security?
 - <http://www.darkreading.com/application-security/database-security/does-nosql-mean-no-security/d/d-id/1136913>

Security ...

- A Response To NoSQL Security Concerns
 - <http://www.darkreading.com/application-security/database-security/a-response-to-nosql-security-concerns/d/d-id/1137044>
- MongoDB - Security Weaknesses in a typical NoSQL database
 - <http://blog.spiderlabs.com/2013/03/mongodb-security-weaknesses-in-a-typical-nosql-database.html>
- Neo4j - “Enter the GraphDB”
 - <http://blog.scr.t.ch/2014/05/09/neo4j-enter-the-graphdb/>

Security

- More Data, More Problems: Part #1
 - <http://blog.imperva.com/2014/08/more-data-more-problems-part-1.html>
- More Data, More Problems: Part #2
 - <http://blog.imperva.com/2014/08/more-data-more-problems-part-2.html>
- More Data, More Problems: Part #3
 - <http://blog.imperva.com/2014/09/more-data-more-problems-part-3.html>

Security alerts ...

- Data, Technologies and Security - Part 1
 - <https://blog.binaryedge.io/2015/08/10/data-technologies-and-security-part-1/>
- Data, Technologies and Security - Part 2
 - <https://blog.binaryedge.io/2016/01/19/data-technologies-and-security-part-1-2/>
- It's the Data, Stupid!
 - <https://blog.shodan.io/its-the-data-stupid/>

Security alerts

- Insecure Data storage with NoSQL Databases
 - <http://resources.infosecinstitute.com/android-hacking-and-security-part-19-insecure-data-storage-with-nosql-databases/>
- MongoDB databases at risk
 - https://cispa.saarland/wp-content/uploads/2015/02/MongoDB_documentation.pdf

NoSQL injection testing ...

- NoSQLMap project
 - <http://nosqlmap.net>
 - <https://github.com/tcstool/NoSQLMap/>
- Making Mongo Cry: NoSQL for Penetration Testers
 - http://www.nosqlmap.net/DC22-WoS-Nosql_slides.pptx

NoSQL injection testing ...

- NoSQL Exploitation Framework
 - <http://nosqlproject.com>
- Pentesting NoSQL DB's with NoSQL Exploitation Framework
 - <https://www.hackinparis.com/node/267/>
 - <http://www.slideshare.net/44Con/pentesting-nosql-dbs-with-nosql-exploitation-framework/>

NoSQL injection testing ...

- Does NoSQL Equal No Injection?
 - <http://securityintelligence.com/does-nosql-equal-no-injection>
- No SQL, No Injection? Examining NoSQL Security
 - <http://arxiv.org/pdf/1506.04082v1>

NoSQL injection testing ...

- Hacking NodeJS and MongoDB
 - <http://blog.websecurify.com/2014/08/hacking-nodejs-and-mongodb.html>
 - <http://java.dzone.com/articles/defending-against-query>
- NoSQL SSJI Authentication Bypass
 - <http://blog.imperva.com/2014/10/nosql-ssji-authentication-bypass.html>

NoSQL injection testing

- Attacking MongoDB
 - <http://www.slideshare.net/cyber-punk/mongo-db-eng/>
- Avoiding MongoDB hash-injection attacks
 - <http://cirw.in/blog/hash-injection>
 - <https://github.com/eoftedal/HashInjection/>
- No SQL injection but NoSQL Injection
 - <http://www.slideshare.net/sth4ck/sthack-2013-florian-agixid-gaultier-no-sql-injection-but-no-sql-injection/>

NoSQL forensics

- NoSQL Forensics: What to do with (No)ARTIFACTS
 - <https://speakerdeck.com/505forensics/nosql-forensics-what-to-do-with-no-artifacts/>
- NoSQL Injections: Moving Beyond or '1'='1'
 - <https://speakerdeck.com/505forensics/nosql-injections-moving-beyond-or-1-equals-1/>
- NoSQL Triage Scripts
 - https://github.com/505Forensics/nosql_triage/

NoSQL honeypot testing

- NoSQL Honeypot Framework (NoPo)
 - <https://github.com/torque59/nosqlpot/>

Polyglot persistence ...

- NoSQL Database Choices: Weather Co. CIO's Advice
 - <http://www.informationweek.com/big-data/software-platforms/nosql-database-choices-weather-co-cios-advice/a/d-id/1317052>
- Why we started using PostgreSQL with Slick next to MongoDB
 - <http://www.plotprojects.com/why-we-use-postgresql-and-slick/>

Polyglot persistence ...

- HBase at Mendeley
 - <http://www.slideshare.net/danharvey/hbase-at-mendeley/>
- Polyglot Persistence
 - <http://www.slideshare.net/jwoodslideshare/polyglot-persistence-two-great-tastes-that-taste-great-together-4625004/>
- Polyglot Persistence Patterns
 - <http://abhishek-tiwari.com/post/polyglot-persistence-patterns/>

Polyglot persistence

- Polyglot Persistence: EclipseLink with MongoDB and Derby
 - <http://java.dzone.com/articles/polyglot-persistence-0>
- D. Ghosh (2010) Multiparadigm data storage for enterprise applications. *IEEE Software*. Vol. 27, No. 5, pp. 57-60

Performance benchmarks ...

- Yahoo Cloud Serving Benchmark
 - <https://github.com/brianfrankcooper/YCSB/>
 - <http://altdoros.com/nosql-research>
 - <http://www.slideshare.net/tazija/evaluating-nosql-performance-time-for-benchmarking/>
 - <http://jaxenter.com/evaluating-nosql-performance-which-database-is-right-for-your-data.1-49428.html>

Performance benchmarks ...

- 2015 YCSB results
 - http://info.couchbase.com/Benchmark_MongoDB_VS_CouchbaseServer_B.html
 - <http://www.mongodb.com/lp/white-paper/benchmark-report/>
 - <http://www.datastax.com/apache-cassandra-leads-nosql-benchmark>

Performance benchmarks ...

- Rising NoSQL Star: Aerospike, Cassandra, Couchbase or Redis?
 - <https://redislabs.com/blog/nosql-performance-aerospike-cassandra-datastax-couchbase-redis>
- Performance comparison between ArangoDB, MongoDB, Neo4j and OrientDB
 - <https://www.arangodb.com/nosql-performance-blog-series/>
 - <https://github.com/weinberger/nosql-tests/>

Performance benchmarks ...

- Performance Evaluation of NoSQL Databases: A Case Study
 - http://www.researchgate.net/publication/275033854_Performance_Evaluation_of_NoSQL_Databases_A_Case_Study
- A Case Study for NoSQL Applications and Performance Benefits: CouchDB vs. Postgres
 - http://figshare.com/articles/A_Case_Study_for_NoSQL_Applications_and_Performance_Benefits_CouchDB_vs_Postgres/787733

Performance benchmarks ...

- Ultra-High Performance NoSQL Benchmarking
 - <http://thumbtack.net/whitepapers/ultra-high-performance-nosql-benchmark.html>
- Comparing NoSQL Data Stores
 - http://www.quantschool.com/home/programming-2/comparing_inmemory_data_stores/
- No SQL Performance Benchmark by SandStorm
 - <http://www.sandstormsolution.com/nosql.html>

Performance benchmarks ...

- NoSQL Performance when Scaling by RAM
 - http://info.couchbase.com/rs/northscale/images/NoSQL_Performance_Scaling_by_RAM.pdf
- Dissecting the NoSQL Benchmark
 - <http://blog.couchbase.com/dissecting-nosql-benchmark/>
- Benchmarking Couchbase Server
 - <http://www.slideshare.net/Couchbase/t1-s4-couchbase-performancebenchmarkingv34/>

Performance benchmarks ...

- NoSQL Performance Benchmarks Series: Couchbase
 - <http://blog.bigstep.com/big-data-performance/nosql-performance-benchmarks-series-couchbase/>
- Benchmarking Riak
 - <https://medium.com/@mustwin/benchmarking-riak-bfee93493419/>

Performance benchmarks ...

- NoSQL Fast? Not always. A benchmark
 - <http://machieltgroeneveld.wordpress.com/2014/07/01/nosql-fast/>
- Finding the right NoSQL data store: Results for my use case and a surprise
 - <https://www.paluch.biz/blog/124-finding-the-right-nosql-data-store-results-for-my-use-case-and-a-surprise.html>

Performance benchmarks ...

- MongoDB Performance Pitfalls - Behind The Scenes
 - <http://blog.trackerbird.com/content/mongodb-performance-pitfalls-behind-the-scenes/>
- MySQL vs. MongoDB Disk Space Usage
 - <http://blog.trackerbird.com/content/mysql-vs-mongodb-disk-space-usage/>
- MongoDB: Scaling write performance
 - <http://www.slideshare.net/daumdna/mongodb-scaling-write-performance/>

Performance benchmarks ...

- MySQL vs MongoDB performance benchmark
 - <http://www.moredevs.com/mysql-vs-mongodb-performance-benchmark/>
- Postgres Outperforms MongoDB and Ushers in New Developer Reality
 - <http://blogs.enterprisedb.com/2014/09/24/postgres-outperforms-mongodb-and-ushers-in-new-developer-reality/>

Performance benchmarks ...

- Can the Elephants Handle the NoSQL Onslaught?
 - http://vldb.org/pvldb/vol5/p1712_avriliafloratou_vldb2012.pdf
- Solving Big Data Challenges for Enterprise Application Performance Management
 - http://vldb.org/pvldb/vol5/p1724_tilmanrabi_vldb2012.pdf
- NoSQL RDF
 - <https://github.com/ahaque/hive-hbase-rdf/>

Performance benchmarks

- Benchmarking Graph Databases
 - <http://istc-bigdata.org/index.php/benchmarking-graph-databases/>
- Benchmarking Graph Databases - Updates
 - <http://istc-bigdata.org/index.php/benchmarking-graph-databases-updates/>
- Linked Data Benchmark Council
 - <http://ldbc.eu/>

Benchmarking tips ...

- How not to benchmark Cassandra
 - <http://www.datastax.com/dev/blog/how-not-to-benchmark-cassandra>
- How not to benchmark Cassandra: a case study
 - <http://www.datastax.com/dev/blog/how-not-to-benchmark-cassandra-a-case-study>
- Scaling NoSQL databases: 5 tips for increasing performance
 - <http://radar.oreilly.com/2014/09/scaling-nosql-databases-5-tips-for-increasing-performance.html>

Benchmarking tips

- How To Benchmark NoSQL Databases
 - <http://blog.bigstep.com/big-data-performance/benchmark-nosql-databases/>
- Correcting YCSB's Coordinated Omission problem
 - <http://psy-lob-saw.blogspot.co.uk/2015/03/fixing-ycsb-coordinated-omission.html>

Jepsen stress testing ...

- Jepsen
 - <http://www.aphyr.com/tags/jepsen>
- Jepsen: Testing the Partition Tolerance of PostgreSQL, Redis, MongoDB and Riak
 - <http://www.infoq.com/articles/jepsen/>
- The Man Who Tortures Databases
 - <http://www.informationweek.com/software/information-management/the-man-who-tortures-databases/240160850/>

Jepsen stress testing ...

- Testing Network failure using NuoDB and Jepsen, part 1
 - <http://dev.nuodb.com/techblog/testing-network-failure-using-nuodb-and-jepsen-part-1>
- Testing Network failure using NuoDB and Jepsen, part 2
 - <http://dev.nuodb.com/techblog/testing-network-failure-using-nuodb-and-jepsen-part-2>

Jepsen stress testing

- Jepsen IV: Hope Springs Eternal
 - <http://www.thedotpost.com/2015/06/kyle-kingsbury-jepsen-iv-hope-springs-eternal>

Unit testing

- Unit Testing NoSQL Databases Applications with NoSQLUnit
 - <http://www.methodsandtools.com/tools/nosqlunit.php>
 - <https://github.com/lordofthejars/nosql-unit/>

BI/Analytics

- BI/Analytics on NoSQL: Review of Architectures Part 1
 - <http://www.dataversity.net/bianalytics-on-nosql-review-of-architectures-part-1/>
- BI/Analytics on NoSQL: Review of Architectures Part 2
 - <http://www.dataversity.net/bianalytics-on-nosql-review-of-architectures-part-2/>

Various graphics ...

- G2 Crowd Grid for NoSQL
 - <https://www.g2crowd.com/categories/nosql-databases/>
- Data Platforms Landscape map
 - <https://451research.com/state-of-the-database-landscape/>
- NoSQL LinkedIn Skills Index - September 2015
 - https://blogs.the451group.com/information_management/2015/10/01/nosql-linkedin-skills-index-september-2015/

Various graphics ...

- Necessity is the mother of NoSQL
 - http://blogs.the451group.com/information_management/2011/04/20/necessity-is-the-mother-of-nosql/
- Making Sense of Big Data
 - <http://www.slideshare.net/infochimps/making-sense-of-big-data/>
- NoSQL, Heroku, and You
 - <https://blog.heroku.com/archives/2010/7/20/nosql/>

Various graphics

- The NoSQL vs. SQL hoopla, another turn of the screw!
 - <http://www.parelastic.com/blog/nosql-vs-sql-hoopla-another-turn-screw/>
- Navigating the Database Universe
 - <http://www.slideshare.net/lisapaglia/navigating-the-database-universe/>

Discussion fora

- LinkedIn NoSQL
 - <http://www.linkedin.com/groups?gid=2085042>
- LinkedIn NewSQL
 - <http://www.linkedin.com/groups/NewSQL-4135938>
- Google groups
 - <http://groups.google.com/group/nosql-discussion>
- Quora
 - <https://www.quora.com/NoSQL/>

NoSQL jokes/humour ...

- LinkedIn discussion thread
 - <http://www.linkedin.com/groups/NoSQL-Jokes-Humour-2085042.S.177321213>
- NoSQL Better Than MySQL?
 - <http://www.youtube.com/watch?v=QU34ZVD2yIY>
 - Shorter version of “Episode 1 - MongoDB is Web Scale”
- /dev/null vs. MongoDB benchmark bake-off
 - <http://engineering.wayfair.com/devnull-vs-mongodb-benchmark-bake-off/>

NoSQL jokes/humour ...

- say No! No! and No! (=NoSQL Parody)
 - <http://www.youtube.com/watch?v=fXc-QDJBXpw>
- **BREAKING: NoSQL just “huge text file and grep”, study finds**
 - <http://thescienceweb.wordpress.com/2014/10/28/breaking-nosql-just-huge-text-file-and-grep-study-finds/>

NoSQL jokes/humour ...

- When someone brags about scaling MongoDB to a whopping 100GB
 - <http://dbareactions.tumblr.com/post/62989609976/when-someone-brags-about-scaling-mongodb-to-a>
- Trying not to use NoSQL when others do
 - <http://devopsreactions.tumblr.com/post/128836122545/trying-not-to-use-nosql-when-others-do>

NoSQL jokes/humour ...

- Interview with the Ghost of MongoDB Scalability
 - <http://blog-shaner.rhcloud.com/interview-with-the-ghost-of-mongodb-scalability/>
- It's Time to Breakup with Your Longtime RDBMS
 - <http://www.marklogic.com/blog/time-breakup-longtime-rdbms/>

NoSQL jokes/humour

- C.R.U.D.
 - <http://crudcomic.tumblr.com/>
- Twitter
 - @mongodbfacts
 - @BigDataBorat

Miscellaneous ...

- PowerPoint template
 - <http://www.articulate.com/rapid-elearning/heres-a-free-powerpoint-template-how-i-made-it/>
- Autostereogram
 - http://www.all-freeware.com/images/full/46590-free_stereogram_screensaver_audio___multimedia_othr.jpeg
- Theatre Curtain Animations
 - <http://www.slideshare.net/chinateacher1/theater-curtain-animations/>

Miscellaneous ...

- Icons and images
 - <http://www.geekpedia.com/icons.php>
 - <http://cemagraphics.deviantart.com/>
 - <http://www.freestockphotos.biz/>
 - <http://www.graphicsfuel.com/2011/09/comments-speech-bubble-icon-psd/>
 - <http://www.softicons.com/free-icons/>
 - <http://icondock.com/>

Miscellaneous

- Newspaper headlines
 - <http://www.imagechef.com/t/n8rm/Newspaper-Headline/>

Backup headlines







Source: Inspired by “BREAKING: NoSQL just ‘huge text file and grep’, study finds” jovialscientist (28 October 2014)